The Shimura-Taniyama conjecture (d'après Wiles)

Henri Darmon

September 9, 2007

Contents

The conjecture of Shimura and Taniyama that every elliptic curve over **Q** is modular has been described as a "Himalayan peak" [Mu] whose conquest is one of the great challenges of mathematics. Until recently, the summit was viewed as impregnable. Then, in June 1993, Andrew Wiles stunned the world by mapping out a means of ascent to those lofty peaks.

Wiles' methods can be used to prove that there are infinitely many elliptic curves over **Q** with distinct j-invariants that are modular. In this survey we will give a fairly complete proof of this result. Our goal here has been to take as many short-cuts as possible, guiding the reader along a beginner's trail which avoids the treacherous slopes of Euler Systems and p-adic Hodge theory, but still leads to a vantage point from which some of Wiles' beautiful achievement can be contemplated.

In the general case, Wiles reduces the Shimura-Taniyama conjecture to a conjectural upper bound on the Selmer groups associated to certain motives of rank 3. He also gives a method for establishing this upper bound, assuming that certain rings of Hecke operators are local complete intersections. The fundamental work of Matthias Flach suggests that a proof of the upper bound might also be possible by using the ideas on "Euler systems" introduced by V.A. Kolyvagin. "Euler system" calculations have been performed before in many different settings, always with spectacular results, most notably by Thaine [Th], Kolyvagin [Ko], Rubin [Ru], Nekovar [Ne], and Flach [Fl].

In any case, Wiles' reduction removes much of the mystery behind the Shimura-Taniyama conjecture – and, to the optimist, suggests that a proof must be within reach! We will very briefly indicate some of the features of Wiles' general strategy at the end.

We do not say anything about the well-known connection between the Shimura-Taniyama conjecture and Fermat's Last Theorem, which is amply documented, for example in the excellent survey articles [Co], [Gou], [Ri4], [RH], [RS].

This paper is entirely expository. It owes everything to the ideas of Wiles, and to a course he and his students gave at Princeton University in the Spring semester 1994. I would also like to thank Brian Conrad, Fred Diamond, Ravi Ramakrishna, Ken Ribet, Anna Rio, Richard Taylor and Larry Washington for making their notes available to me, answering my questions, and making comments and corrections on earlier versions of this manuscript. Some notes of H.W. Lenstra [Le] on the ring theoretic aspects of Wiles work were also tremendously helpful in writing chapter 2.

1 Preliminaries

1.1 The main results

We say that $j \in \mathbf{Q}$ is *modular* if it is the j-invariant of a modular elliptic curve over $\mathbf{Q}$. It follows from the work of Shimura that if j is modular, then *every* elliptic curve with invariant j is modular.

Theorem 1.1 *(Wiles) There are infinitely many $j \in \mathbf{Q}$ which are modular.*

This result represents a major watershed: before Wiles' announcement, a proof seemed to be nowhere in sight.

If E is an elliptic curve, then we denote by E_p the group of p-division points of E; it is a two-dimensional $\mathbf{F}_p$-vector space, endowed with a natural action of $G_{\mathbf{Q}} = \mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$.

The infinite set of modular j-invariants can be obtained via the following more precise statement:

Theorem 1.2 *(Wiles) Let X be a semistable modular elliptic curve, let d be the degree of a minimal modular parametrization $X_0(N) \longrightarrow X$, and let p be an odd prime such that p does not divide d and X_p is absolutely irreducible as a Galois module.*

If E is an elliptic curve over $\mathbf{Q}$ such that

$$E_p \simeq X_p \qquad \textit{as Galois modules},$$

and E has good or semistable reduction at p, then the curve E is modular.

Remarks:
1. There are many examples of pairs (X, p) which satisfy the assumptions of the theorem, for example:

$X = X_0(11)$, p any odd prime $\neq 5$,
$X = X_0(14), p > 3$,
$X = X_0(15)$ or $X_0(17)$, p any odd prime, etc...

2. Let $X(p)_{/\mathbf{Q}}$ be the modular curve over $\mathbf{Q}$ which classifies pairs (E, ϕ) where E is an elliptic curve and ϕ is a monomorphism $\phi : \mathbf{Z}/p\mathbf{Z} \times \mathbf{Z}/p\mathbf{Z} \longrightarrow E$. The group scheme X_p can be used to define the twisted modular curve $X'(p)_{/\mathbf{Q}}$, which classifies pairs (E, ϕ) with $\phi : X_p \hookrightarrow E$ is an injection. When $p = 3$ or $p = 5$, the curves $X(p)$ and $X'(p)$ are a union of $p - 1$ curves of genus 0.

(For example, the curve $X(5)$ is related to the classical "icosahedral curve" studied by Klein.) The pair $(X, \phi : X_p \hookrightarrow X)$ then defines a rational point x on one of the components of $X^{'}(p)_{/\mathbf{Q}}$. This component is a curve of genus 0 having a $\mathbf{Q}$-rational point, and hence is isomorphic to $\mathbf{P}_{1/\mathbf{Q}}$. The rational points in $X^{'}(p)(\mathbf{Q})$ which are close to x in the p-adic topology, give rise to an infinite set of elliptic curves E satisfying the hypothesis of the theorem. (In fact, they yield a family of modular j-invariants which are the values at certain rational arguments of a rational function of degree 12 when $p = 3$, and 60 when $p = 5$.) For $p > 5$, thm. 1.2 is not as interesting, since the curves E satisfying the hypothesis for a fixed X correspond to rational points on a twist of $X(p)$, which has genus strictly greater than 1; by Faltings' proof of the Mordell conjecture, there are only finitely many such E.

We will explain the proof of theorem 1.2 only in the special case where $X = X_0(17)$ and $p = 5$. This allows us to lighten the exposition by avoiding a number of technical issues, while still being sufficient to prove thm. 1.1.

Theorem 1.3 *Let E be an elliptic curve over $\mathbf{Q}$ which has good reduction at 5 and satisfies*

$$E_5 \simeq X_0(17)_5 \quad \textit{as Galois modules.}$$

Then E is modular.

Remarks:
1. The curve $X = X_0(17)$ is a curve of genus 1, and hence is a modular elliptic curve; the degree d of the modular parametrization is 1. Also, as we will see in lemma 1.4, the module X_5 gives a two-dimensional mod 5 representation of $G_{\mathbf{Q}} = \mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$ which is irreducible, and in fact, surjective. Hence, thm. 1.3 is a special case of thm. 1.2.
2. The statement of thm. 1.3 would also be true with 5 replaced by 3; working with 5 instead of 3 allows us to avoid certain slight technical difficulties. Note, however, that the prime 3 plays a crucial role in Wiles' general strategy for proving the Shimura-Taniyama conjecture; cf. sec. 5.

From now on, let $\bar{\rho}_0$ be the mod 5 representation arising from the 5 division points of $X_0(17)$,

$$\bar{\rho}_0 : G_{\mathbf{Q}} \longrightarrow \mathrm{Aut}\,(X_5) \simeq \mathbf{GL}_2(\mathbf{F}_5).$$

Lemma 1.4 *The representation $\bar{\rho}_0$ satisfies the following properties:*

1. *The character* $\det(\bar{\rho}_0)$ *is the cyclotomic character* $\bar{\epsilon} : G_{\mathbf{Q}} \longrightarrow \mathbf{F}_5^*$ *giving the action of* $G_{\mathbf{Q}}$ *on the* 5*th roots of unity.*

2. *(Local behaviour at* 5*). The decomposition group* D_5 *at* 5 *acts on* X_5 *by preserving a one-dimensional subspace, and the inertia group* I_5 *acts trivially on the quotient:*

$$\bar{\rho}_0|_{D_5} \simeq \begin{pmatrix} \bar{\chi}_1 & * \\ 0 & \bar{\chi}_2 \end{pmatrix}, \quad \bar{\rho}_0|_{I_5} \simeq \begin{pmatrix} \bar{\epsilon} & * \\ 0 & 1 \end{pmatrix}.$$

Furthermore, the unramified character $\bar{\chi}_2$ *is of order* 4.

3. *(Local behaviour at* 17*):*

$$\bar{\rho}_0|_{D_{17}} \simeq \begin{pmatrix} \bar{\epsilon} & \bar{\Psi} \\ 0 & 1 \end{pmatrix},$$

and $\bar{\Psi}|_{I_{17}}$ *is non-trivial.*

4. *The representation* $\bar{\rho}_0$ *is absolutely irreducible - in fact, it is surjective.*

Proof:
1. The Weil pairing gives a $G_{\mathbf{Q}}$-equivariant isomorphism $\wedge^2 X_5 = \mu_5$, and hence $\det(\bar{\rho}_0) = \bar{\epsilon}$.
2. For each prime l, let $a_l = l + 1 - \#X(\mathbf{F}_l)$. Since $a_5 = -2 \neq 0$, the curve X has good *ordinary* reduction at 5. Hence $X_5(\bar{\mathbf{F}}_5) \simeq \mathbf{Z}/5\mathbf{Z}$, and there is an exact sequence of I_5-modules

$$0 \longrightarrow \mu_5 \longrightarrow X_5 \longrightarrow \mathbf{Z}/5\mathbf{Z} \longrightarrow 0$$

induced by the reduction map $X(\bar{\mathbf{Q}}_5) \longrightarrow X(\bar{\mathbf{F}}_5)$. Let $W = X_5(\bar{\mathbf{F}}_5) \simeq \mathbf{Z}/5\mathbf{Z}$ be the unramified quotient of X_5. Since

$$\#X(\mathbf{F}_5) = 5 + 1 - a_5 = 8, \quad \#X(\mathbf{F}_{25}) = 25 + 1 - a_{25} = 32,$$

the character $\bar{\chi}_2$ cannot be of order 1 or 2; hence it is of order 4.
3. The curve X has split multiplicative reduction at 17, and is isomorphic over $\mathbf{Q}_{17}$ to a Tate curve $G_m/q^{\langle \mathbf{Z} \rangle}$ where $q \in 17\mathbf{Z}_{17}$ is the Tate parameter.

From this explicit model, one sees that the representation $\bar{\rho}_0|_{I_{17}}$ is as given in the proposition. To see that $\bar{\Psi}$ is non-trivial, let Δ be the minimal discriminant of X at 17. One can check from tables [Cr] that $\mathrm{ord}_{17}(\Delta) \neq 0 \pmod 5$. Hence

$$\mathrm{ord}_{17}(q) = -\mathrm{ord}_{17}(\Delta) \neq 0 \pmod 5,$$

so that $\bar{\Psi}|_{I_{17}} \neq 0$.
4. Part 4 follows from prop. 21, p. 306, of [Sr1], using the fact that

$$\mathrm{ord}_{17}(\Delta) \neq 0 \pmod 5 \text{ and } 2+1-a_2 \neq 0 \pmod 5.$$

A deformation of $\bar{\rho}_0$ is an equivalence class of representations

$$\rho : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(A)$$

where A is a complete Noetherian local $\mathbf{Z}_5$-algebra with residue field $\mathbf{F}_5$, such that the residual representation

$$G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(A) \longrightarrow \mathbf{GL}_2(\mathbf{F}_5)$$

is equal to $\bar{\rho}_0$. These representations are taken modulo conjugation by matrices in $\mathbf{GL}_2(A)$ which reduce to the identity in $\mathbf{GL}_2(\mathbf{F}_5)$.

The obvious example of a deformation of $\bar{\rho}_0$ is ρ_0, the 5-adic representation associated to the 5-adic Tate module of $X_0(17)$.

We now define the two interesting classes of deformations that we wish to compare.

Definition 1.5 *A deformation $\rho : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(A)$ of $\bar{\rho}_0$ is* admissible *if*

1. *The character* $\det(\rho)$ *is the cyclotomic character $\epsilon : G_{\mathbf{Q}} \longrightarrow \mathbf{Z}_5^* \subset A^*$ giving the action of $G_{\mathbf{Q}}$ on the 5^nth roots of unity.*

2. *(Local behaviour at 5). The decomposition group D_5 at 5 acts on the underlying rank 2 A-module by preserving a free submodule of rank 1 with free quotient, and the inertia group I_5 acts trivially on this quotient:*
$$\rho|_{D_5} \simeq \begin{pmatrix} \chi_1 & * \\ 0 & \chi_2 \end{pmatrix}, \quad \rho|_{I_5} \simeq \begin{pmatrix} \epsilon & * \\ 0 & 1 \end{pmatrix}.$$

3. (Local behaviour at 17*):*

$$\rho|_{D_{17}} \simeq \begin{pmatrix} \epsilon & \Psi \\ 0 & 1 \end{pmatrix}.$$

Part of the motivation for this definition, in the context of our problem, is suggested by the following proposition:

Proposition 1.6 *If E is an elliptic curve satisfying the assumption of thm. 1.3, then the 5-adic representation ρ_E associated to E is admissible. (In particular, ρ_0 is an admissible deformation of $\bar{\rho}_0$.)*

Proof: We check that the representation ρ_E of $G_{\mathbf{Q}}$ acting on the 5-adic Tate module $Ta_5(E)$ of E satisfies the three properties in the definition of admissibility.
1. The Weil pairing gives a $G_{\mathbf{Q}}$-equivariant isomorphism $\wedge^2 Ta_5(E) = \mathbf{Z}_5(1)$, and hence $\det(\rho_E) = \epsilon$.
2. If E has good reduction at 5, the assumption that $\bar{\rho}_E \simeq \bar{\rho}_0$ forces E to have good *ordinary* reduction at 5, since $X_0(17)$ does. For, otherwise the image of the inertia group I_5 at 5 under $\bar{\rho}_E$ would be a non-split cartan subgroup, (cf. [Sr1], prop. 12, p. 275), and $\bar{\rho}_E$ could not be isomorphic to $\bar{\rho}_0$. By the ordinariness of E, there is an exact sequence of $\mathbf{Z}_5[I_5]$-modules

$$0 \longrightarrow \mathbf{Z}_5(1) \longrightarrow Ta_5(E) \longrightarrow \mathbf{Z}_5 \longrightarrow 0,$$

where the rightmost map comes from the reduction map $E(\bar{\mathbf{Q}}_5) \longrightarrow E(\bar{\mathbf{F}}_5)$. If E has multiplicative reduction, then, after twisting by an unramified character, the curve $E_{/\mathbf{Q}_5}$ becomes isomorphic to a Tate curve $G_m/q^{\langle \mathbf{Z} \rangle}$, where $q \in 5\mathbf{Z}_5$. This implies that the representation $\rho_E|_{D_5}$ is of the form

$$\rho_E|_{D_5} \simeq \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}$$

where β is a character of D_5/I_5 of order 1 or 2. By part 2 of lemma 1.4, one has $\bar{\rho}_E \neq \bar{\rho}_0$, and hence, the case where E has multiplicative reduction is ruled out.
3. The proof that the local condition at 17 is satisfied proceeds along similar lines. One notes that E necessarily has multiplicative reduction at 17, and then uses the Tate model to analyze the behaviour of ρ_E restricted to D_{17}.

The second class of deformations that we wish to consider are the *modular deformations.* We say that f is a cusp form of weight 2 (or simply a cusp form, since we will deal only with weight 2) if it is holomorphic on the extended upper half plane, vanishes at the cusps, and satisfies the usual transformation property under a congruence subgroup $\Gamma_0(N)$ for some N. By the q-expansion principle, the space $S_2(N, \mathbf{Z})$ of such cusp forms having integral Fourier expansions is a free $\mathbf{Z}$-module of rank $g = \text{genus}(X_0(N))$, and

$$S_2(N, \mathbf{Z}) \otimes \mathbf{C} = S_2(N, \mathbf{C}),$$

the space on the right being the usual space of holomorphic cusp forms. This fact allows us to define the space $S_2(N, A)$ of cusp forms with values in a ring A by

$$S_2(N, A) := S_2(N, \mathbf{Z}) \otimes A.$$

The module $S_2(N, \mathbf{Z})$ is endowed with an action of the Hecke operators T_p for $(p, N) = 1$ and U_q for $q|N$. This action can be extended to $S_2(N, A)$ by linearity.

An *eigenform* (with coefficients in A) is a form in $S_2(N, A)$ which is an eigenfunction for all the Hecke operators T_p and U_q. We always assume that eigenforms are normalized so that their first Fourier coefficient is 1, so that f can be expanded about the cusp ∞ as

$$f = \sum_{n=1}^{\infty} a_n q^n, \quad a_1 = 1, \qquad q = e^{2\pi i \tau}.$$

Let $\mathcal{O}_f$ be the subring of A generated by the $a_p, p \not| N$. (Note that in general, $\mathcal{O}_f$ may be strictly contained in the ring generated by all the Fourier coefficients of f.)

Definition 1.7 *A deformation $\rho : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(A)$ of $\bar{\rho}_0$ is said to be* modular *if there exists a normalized eigenform $f = \sum_n a_n q^n$ of weight 2 and some level N, with coefficients in $A' \supset A$, such that the element $\rho(\text{Frob}_l)$ has characteristic polynomial*

$$x^2 - a_l x + l,$$

for all l not dividing $5N$. Futhermore, a modular deformation of $\bar{\rho}_0$ is said to be good *if* 5 *does not divide N.*

Proposition 1.8 *Every good modular deformation of $\bar{\rho}_0$ is admissible.*

This follows from the construction, due to Eichler and Shimura, of the representations attached to normalized cusp forms of weight 2 (cf. [Sh]), and the work of Carayol which, among other things, analyzes their basic properties at the bad places (cf. [Ca1]).

What Wiles shows is the converse of this statement.

Theorem 1.9 *(Wiles) Every admissible deformation of $\bar{\rho}_0$ is modular.*

This theorem implies theorem 1.3. For, if E is an elliptic curve satisfying the assumption of thm. 1.3, then the Galois representation on the 5-adic Tate module of E gives an admissible representation

$$\rho_E : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(\mathbf{Z}_5)$$

which is a deformation of $\bar{\rho}_0$. Hence it is modular. It follows from the isogeny conjectures proved by Faltings that E appears in (is isogenous to a factor of) the jacobian of a modular curve, and hence, E itself is modular.

We will now concentrate on proving thm. 1.9.

1.2 Wiles' strategy

Wiles' proof of thm. 1.9 is a sophisticated counting argument. Roughly speaking, Wiles shows that the map

$$\{\text{Modular deformations of } \bar{\rho}_0\} \longrightarrow \{\text{Admissible deformations of } \bar{\rho}_0\}$$

is an isomorphism (of sets) by counting the respective orders of these sets. This is not quite true, because both of these sets are infinite. To cut down their sizes, one fixes a finite set of primes, Σ, which does not contain the "exceptional set" $\{5, 17\}$. One says that a deformation of $\bar{\rho}_0$ is Σ-admissible (Σ-modular) if it is admissible (modular) and is unramified outside $\Sigma \cup \{5, 17\}$.

One now considers the map

$$\{\Sigma\text{-Modular deformations of } \bar{\rho}_0\} \longrightarrow \{\Sigma\text{-Admissible deformations of } \bar{\rho}_0\}$$

and attempts to show it is an isomorphism; the set on the right is defined purely in terms of the properties of $G_\Sigma = \mathrm{Gal}(\mathbf{Q}_\Sigma/\mathbf{Q})$, the Galois group of the maximal extension unramified outside $\Sigma \cup \{5, 17\}$. As we will see later, the size of this rather subtle set is controlled by a subgroup of a Galois cohomology group – the Selmer group of the symmetric square associated to ρ_0. Wiles' strategy is to bound the size of this Selmer group in terms of "modular" data related to the set on the left, and thus show that the set on the right cannot be bigger than the set on the left, so that all the admissible deformations are forced to be modular.

However, the sets above, even with the restrictions associated to Σ, are still infinite. To obtain finite sets one needs to fix a $\mathbf{Z}_5$-algebra, A, and restrict one's attention to the (equivalence classes of) deformations of $\bar{\rho}_0$ with values in A. A better-posed problem is thus to show that the map

$$\left\{\begin{array}{l}\Sigma\text{-Modular deformations}\\ \text{ of } \bar{\rho}_0\\ \text{with values in } A\end{array}\right\} \longrightarrow \left\{\begin{array}{l}\Sigma\text{-Admissible deformations}\\ \text{ of } \bar{\rho}_0\\ \text{with values in } A\end{array}\right\}$$

is an isomorphism for all local Noetherian $\mathbf{Z}_5$-algebras A. Let $MD_\Sigma(A)$ and $AD_\Sigma(A)$ denote these two sets. The assignments $A \mapsto MD_\Sigma(A)$ and $A \mapsto AD_\Sigma(A)$ are functors from the category of local $\mathbf{Z}_5$-algebras to sets. With these new notions, one can reformulate the problem as that of showing that the "natural" natural transformation

$$MD_\Sigma \longrightarrow AD_\Sigma$$

is a *natural equivalence of functors.*

A key fact is that the functor AD_Σ is *representable*, i.e., there exists a local Noetherian $\mathbf{Z}_5$-algebra R_Σ such that

$$AD_\Sigma = \mathrm{Spec}(R_\Sigma).$$

(Here we are identifying a scheme with its functor of points.)

From the work of Ribet and others, one also expects that the modular deformation functor MD_Σ should be representable by a finite flat local $\mathbf{Z}_5$-algebra $\mathbf{T}_\Sigma$ constructed by completing an appropriate ring of Hecke operators acting on $S_2(N_\Sigma, \mathbf{Z})$, for some level N_Σ depending on Σ.

The natural transformation $MD_\Sigma \longrightarrow AD_\Sigma$ would then translate into a ring homomorphism

$$\phi : R_\Sigma \longrightarrow \mathbf{T}_\Sigma$$

which can be studied by using the machinery of commutative algebra.

1.3 Representability of AD_Σ and the ring R_Σ

In this section we state the result on the representability of AD_Σ.

Theorem 1.10 *(Mazur): The functor of Σ-admissible deformations of $\bar{\rho}_0$ is equal to* $\hom(R_\Sigma, -)$*, where R_Σ is a finitely generated local $\mathbf{Z}_5$-algebra with residue field $\mathbf{F}_5$.*

The ring R_Σ is called the *universal deformation ring* associated to the representation $\bar{\rho}_0$ and the set Σ. Its construction is based on a formal representability criterion of Schlessinger, and is described in [Mz2] (or, see also [MT]).
Remark: A more general result of Ramkrishna [Ra] establishes the existence of the universal deformation ring in certain cases where the residual representation is not ordinary. We will not need this for our application, however.

By definition, there is a "universal admissible deformation"

$$\rho_{\Sigma,univ} : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(R_\Sigma)$$

such that for any Σ-admissible deformation $\rho : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(A)$ of $\bar{\rho}_0$, there is a unique homomorphism $\alpha : R_\Sigma \longrightarrow A$ such that $\rho_{\Sigma,univ} \otimes_\alpha A = \rho$. In particular, the deformation ρ_0 corresponding to $Ta_5(X_0(17))$ gives rise to a natural "base point map"

$$\pi_{R_\Sigma} : R_\Sigma \longrightarrow \mathbf{Z}_5,$$

which will play an important role later on.

1.4 Construction of a Hecke ring $\mathbf{T}_\Sigma$

We are unable to establish a priori that the modular deformation functor is representable. However, we do have:

Expectation 1.11 *The set of good Σ-modular deformations of $\bar{\rho}_0$ is equal to* $\hom(\mathbf{T}_\Sigma, -)$*, where $\mathbf{T}_\Sigma$ is a finite flat local $\mathbf{Z}_5$-algebra with residue field $\mathbf{F}_5$.*

In our situation, Wiles proposes a precise construction of $\mathbf{T}_\Sigma$. We briefly describe this construction here. Let

$$N_\Sigma = 17 \prod_{p \in \Sigma} p^2,$$

let $\Gamma_0(N_\Sigma)$ be the usual congruence subgroup of level N_Σ, and let

$$X_0(N_\Sigma) = \Gamma_0(N_\Sigma)\backslash\mathcal{H} \cup \{\text{cusps}\}$$

be the corresponding curve, compactified in the usual way by adjoining the cusps.

Now let $\mathbf{T}(\Sigma)$ be the ring of Hecke operators acting on $S_2(N_\Sigma, \mathbf{Z})$, the space of cusp forms of weight 2 on $X_0(N_\Sigma)$ with integer Fourier coefficients. This ring is generated over $\mathbf{Z}$ by the Hecke operators T_l, for l not diving N_Σ, and by the Hecke operators U_q for $q \in \Sigma \cup \{17\}$.

Let

$$f = \sum_n a_n q^n$$

be the normalized eigenform of level 17 corresponding to $X = X_0(17)$. Its L-function is equal to

$$L(f,s) = \sum_n a_n n^{-s} = (1 - 17^{-s})^{-1} \prod_{p \neq 17} (1 - a_p p^{-s} + p^{1-2s})^{-1}.$$

The form f is an eigenform for all the Hecke operators in $\mathbf{T}(\emptyset)$ acting on $S_2(17, \mathbf{Z})$, but not for $\mathbf{T}(\Sigma)$ when viewed as a form in $S_2(N_\Sigma, \mathbf{Z})$, since it is no longer an eigenform for the Hecke operators U_q, with $q \in \Sigma$. We remedy this problem by defining a cusp form f_Σ on $X_0(N_\Sigma)$ which is an eigenform for $\mathbf{T}(\Sigma)$ by the inductive formula:

$$f_\emptyset = f, \quad f_{\Sigma\cup\{q\}} = f_\Sigma(\tau) - a_q f_\Sigma(q\tau) + q f_\Sigma(q^2\tau).$$

The reader can check that f_Σ is a modular form on $\Gamma_0(N_\Sigma)$ which satisfies

$$T_l f_\Sigma = a_l f_\Sigma \ (l \notin \Sigma \cup \{17\}), \quad U_q f_\Sigma = 0 \ (q \in \Sigma), \quad U_{17} f_\Sigma = f_\Sigma,$$

and hence is an eigenform for all of $\mathbf{T}(\Sigma)$. The construction of f_Σ corresponds to “removing the Euler factors at the primes of Σ”, since:

$$L(f_\Sigma, s) = (1 - 17^{-s})^{-1} \prod_{p \notin \Sigma \cup \{17\}} (1 - a_p p^{-s} + p^{1-2s})^{-1}.$$

It follows that the ideal

$$\mathcal{M}_\Sigma = \langle 5, T_l - a_l, U_q, U_{17} + 1 \rangle \subset \mathbf{T}(\Sigma)$$

is a proper ideal of $\mathbf{T}(\Sigma)$. Let $\mathbf{T}_\Sigma$ denote the completion of $\mathbf{T}(\Sigma)$ at $\mathcal{M}_\Sigma$,

$$\mathbf{T}_\Sigma = \varprojlim \mathbf{T}(\Sigma)/\mathcal{M}_\Sigma^n.$$

The ring $\mathbf{T}_\Sigma$ is a local $\mathbf{Z}_5$-algebra with residue field $\mathbf{F}_5$. It is finite and flat (i.e., finitely generated and free as a $\mathbf{Z}_5$-module). If $\mathcal{O}$ is any $\mathbf{Z}_5$-algebra with residue field $\mathbf{F}_5$, the $\mathbf{Z}_5$-algebra homomorphisms $\mathbf{T}_\Sigma \longrightarrow \mathcal{O}$ correspond to 5-adic eigenforms on $\Gamma_0(N_\Sigma)$ which are congruent to f_Σ mod 5. In particular, the form f_Σ defines a canonical homomorphism

$$\pi_{T_\Sigma} : \mathbf{T}_\Sigma \longrightarrow \mathbf{Z}_5.$$

The fundamental construction of Eichler and Shimura, which associates a Galois representation to a cusp form of weight 2, completed by the work of Carayol, gives the following theorem:

Theorem 1.12 *There is a (unique, up to conjugation) Galois representation*

$$\rho_{\Sigma,mod} : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(\mathbf{T}_\Sigma)$$

which is admissible, and satisfies

$$trace(\rho_{mod}(\mathrm{Frob}_l)) = T_l, \quad \det(\rho_{mod}(\mathrm{Frob}_l)) = l, \quad \forall l \notin \Sigma \cup \{5, 17\}.$$

Furthermore, this representation is admissible, and the operator U_{17} is the eigenvalue of Frob$_{17}$ *acting on the one-dimensional unramified quotient at 17.*

Proof: See for example [Sh], [Ca1].
Remark: It is not clear a priori that the ring $\mathbf{T}_\Sigma$ represents the functor MD_Σ as we have defined it, although such a result is suggested by the work of Ribet [Ri2], and it will eventually follow a posteriori from the proof of thm. 1.9. A crucial ingredient in Wiles' proof of thm. 1.3 is to first establish this for certain sets Σ.

Naturally, the representability of MD_Σ as it was defined naively before is not logically necessary for Wiles' argument, so that our argument is not circular. One just takes a different tack, and *defines* the functor MD_Σ to be $\hom(\mathbf{T}_\Sigma, -)$.

1.5 The map $R_\Sigma \longrightarrow \mathbf{T}_\Sigma$

By thm. 1.12 and the universality property of the universal deformation ring R_Σ, there is a canonical ring homomorphism

$$\phi_\Sigma : R_\Sigma \longrightarrow \mathbf{T}_\Sigma.$$

Moreover, this homomorphism is compatible with the base point maps, i.e., the diagram

$$\begin{tikzcd} R_\Sigma \arrow[r] \arrow[d] & \mathbf{T}_\Sigma \arrow[d] \\ \mathbf{Z}_5 \arrow[r, phantom, "="] & \mathbf{Z}_5 \end{tikzcd}$$

commutes, where the vertical arrows are the base point maps π_{R_Σ} and π_{T_Σ}. For, it follows directly from the Chebotarev density theorem that the maps π_{R_Σ} and $\pi_{T_\Sigma} \circ \phi_\Sigma$ induce equivalent (i.e., equal) deformations, so that these maps are the same, by the universality property of R_Σ.

Now, Wiles' theorem 1.9 can be restated (yet again) as

Theorem 1.13 (Wiles) *The map $\phi_\Sigma : R_\Sigma \longrightarrow \mathbf{T}_\Sigma$ is an isomorphism.*

From now on we will concentrate on the proof of thm. 1.13.

In this section we will show that ϕ_Σ is surjective, which is (by far!) the easy half of thm. 1.13.

Theorem 1.14 *The map ϕ_Σ is surjective.*

Let $\mathbf{T}^0(\Sigma) \subset \mathbf{T}(\Sigma)$ be the subring generated by the "good" Hecke operators T_l, for $l \notin \{5, 17\} \cup \Sigma$, and let $\mathbf{T}^0_\Sigma$ be the closure of the image of $\mathbf{T}^0(\Sigma)$ in $\mathbf{T}_\Sigma$.

Lemma 1.15 *The image of ϕ_Σ contains $\mathbf{T}^0_\Sigma$.*

Proof: For all $l \notin \{5, 17\} \cup \Sigma$,

$$T_l = \mathrm{Tr}\,(\rho_{\Sigma,mod}(\mathrm{Frob}_l)) = \phi_\Sigma(\mathrm{Tr}\,(\rho_{\Sigma,univ}(\mathrm{Frob}_l))),$$

so that $T_l \in \phi_\Sigma(R_\Sigma)$, for all $l \notin \{5, 17\} \cup \Sigma$. The result follows.

Lemma 1.16 *If $q \in \Sigma$, then the image of the Hecke operator U_q in $\mathbf{T}_\Sigma$ is 0. In particular, $U_q \in \mathbf{T}^0_\Sigma$.*

Proof: The operator U_q in $\mathbf{T}(\Sigma)$ satisfies the relation:

$$U_q(U_q^2 - 1)\prod_g (U_q^2 - a_q(g)U_q + q) = 0$$

where the product is taken over all the newforms of level dividing N_Σ/q^2 (cf. [Kn], prop. 9.27, p.289). Since U_q belongs to $\mathcal{M}$, its image in $\mathbf{T}_\Sigma$ is topologically nilpotent. It follows that the expression $(U_q^2 - 1)\prod(U_q^2 - a_q(g)U_q + q)$ maps to a unit in $\mathbf{T}_\Sigma$, so that the image of U_q in $\mathbf{T}_\Sigma$ is 0. The result follows.

To see that the operators U_{17} and T_5 belong to $\phi(R_\Sigma)$ requires a better understanding of $\rho_{\Sigma,mod}$.

Lemma 1.17 *The representation $\rho_{\Sigma,mod} : G_{\mathbf{Q}} \longrightarrow \mathbf{GL}_2(\mathbf{T}_\Sigma)$ is conjugate to a representation with values in $\mathbf{GL}_2(\mathbf{T}_\Sigma^0)$.*

Proof: This follows directly from the following more general statement. (I am grateful to J-P. Serre for explaining this to me.)

Proposition 1.18 *Let $A^0 \subset A$ be an inclusion of local rings (i.e., if m_A and m_{A^0} are the maximal ideals of A and A^0, then $m_{A^0} = m_A \cap A^0$, and A^0 and A have the same residue fields), and suppose that the residue field k of A has trivial Brauer group, $H^2(k, \bar{k}^*) = 0$. Let $\rho : G \longrightarrow \mathbf{GL}_n(A)$ be a representation of a group G over A, and suppose that*
1. The residual representation $\bar{\rho} : G \longrightarrow \mathbf{GL}_n(k)$ is absolutely irreducible.
2. Tr$(\rho(\sigma)) \in A^0$ for all $\sigma \in G$.
Then ρ is conjugate to a representation with values in $\mathbf{GL}_n(A^0)$.

Proof: Let B be the A^0-subalgebra of $M_n(A)$ generated by $\rho(G)$. Since $\bar{\rho}$ is absolutely irreducible, the image of B in $M_n(k)$ is an Azumaya algebra over k, and hence is the full matrix ring $M_n(k)$ since $H^2(k, \bar{k}^*) = 0$. Let $e_1, \ldots, e_{n^2}$ be elements of B that reduce to a standard basis of $M_n(k)$. We claim that $e_1, \ldots, e_{n^2}$ is an A^0 basis for B. By Nakayama's lemma, it is an A-basis for $M_n(A)$, and hence every element b of B can be written uniquely as a combination

$$b = \sum a_i e_i, \qquad \text{with } a_i \in A.$$

We claim that the a_i actually belong to A^0. For, multiplying the above relation by the transpose e_j^t of e_j and taking traces, we find the system:

$$\mathrm{Tr}\,(b e_j^t) = \sum a_i \mathrm{Tr}\,(e_i e_j^t), \quad j = 1, \ldots, n^2. \qquad (1)$$

The matrix $(c_{ij}) = (\mathrm{Tr}\,(e_i e_j^t))$ is an $n^2 \times n^2$-matrix with coefficients in A^0, by assumption. Since the e_i reduce to the standard basis of $M_n(k)$, the image of (c_{ij}) in $M_{n^2}(k)$ is the identity matrix. Hence, (c_{ij}) is invertible in $M_{n^2}(A^0)$, so that the system (1) can be solved for the a_i, and hence $a_i \in A^0$. It follows that B is an algebra of rank n^2 over A^0, generated by the e_i. Let $V \subset A^n$ be the A^0-module generated by the first columns (say) of elements of B. Then V is a free A^0-module of rank n, and the action of B on V gives a map $B \longrightarrow \mathrm{end}(V) \simeq M_n(A^0)$. By Nakayama $B \simeq M_n(A^0)$, and the result follows.

Lemma 1.19 *The image of* U_{17} *in* $\mathbf{T}_\Sigma$ *belongs to* $\mathbf{T}^0_\Sigma$.

Proof: By lemma 1.17, the representation $\rho_{\Sigma,mod}$ is conjugate to a representation with values in $\mathbf{GL}_2(\mathbf{T}^0_\Sigma)$. But U_{17} can be recovered as the eigenvalue of Frob_{17} acting on the unique unramified one-dimensional quotient of the underlying $\mathbf{T}^0_\Sigma$-module, by thm. 1.12. Therefore, $U_{17} \in \mathbf{T}^0_\Sigma$.

Lemma 1.20 *The image of* T_5 *in* $\mathbf{T}_\Sigma$ *belongs to* $\mathbf{T}^0_\Sigma$.

Proof: One recovers T_5 as $\alpha_5 + 5/\alpha_5$, where $\alpha_5 \in \mathbf{T}^0_\Sigma$ is the eigenvalue of Frob_5 acting on the unique unramified rank 1 quotient of $\rho_{\Sigma,mod}$, by thm. 1.12. Hence, T_5 belongs to $\mathbf{T}^0_\Sigma$, as before.

Corollary 1.21 *The natural inclusion* $\mathbf{T}^0_\Sigma \subset \mathbf{T}_\Sigma$ *is an isomorphism.*

Proof: This follows immediately from lemmas 1.16, 1.19, and 1.20.

Proof of thm. 1.14: Combine lemma 1.15 and cor. 1.21.

2 Wiles' isomorphism criterion

Wiles gives a beautiful criterion, based entirely on commutative algebra, to show that certain surjective maps of local rings are isomorphisms.

As we saw in the previous section, the rings that come up in Wiles' situation are finitely generated complete local $\mathbf{Z}_p$-algebras A and are equipped with a natural surjective map $A \longrightarrow \mathbf{Z}_p$. Thus it is natural to work in the category $\mathcal{C}$ whose objects are pairs (A, π), where A is a finitely generated local $\mathbf{Z}_p$-algebra and $\pi : A \longrightarrow \mathbf{Z}_p$ is a surjective ring homomorphism (called the "base point map"). Morphisms in this category are local ring homomorphisms which are compatible in the obvious way with the base point maps. By abuse of notation we often omit mentioning the base point map π when we talk of objects in $\mathcal{C}$, and simply use A to denote (A, π), when this causes no confusion.

2.1 The invariants Φ and η

2.1.1 Definition

Let $A = (A, \pi)$ be an object of $\mathcal{C}$. One associates to such an object two fundamental invariants:

$$\begin{aligned} \Phi_A &= (\ker \pi)/(\ker \pi)^2; \qquad (2) \\ \eta_A &= \pi(\mathrm{Ann}_A \ker \pi). \qquad (3) \end{aligned}$$

The invariant Φ_A can be thought of as a tangent space for the object A. (More precisely, it is the cotangent space for the scheme spec(A) at the point $\ker \pi$.) It is a finitely generated $\mathbf{Z}_p$ module.

The invariant η_A is called the congruence ideal. The reason for this terminology should become clear in section 2.1.2.

We can now state Wiles' isomorphism criterion:

Theorem 2.1 *(Wiles) Let R and T be objects of $\mathcal{C}$ such that T is a finitely generated torsion-free $\mathbf{Z}_p$-module, and let $\phi : R \longrightarrow T$ be a surjective morphism. If*

$$\#\Phi_R \leq \#(\mathbf{Z}_p/\eta_T) < \infty,$$

then ϕ is an isomorphism.

This beautiful result is the engine which is at the heart of Wiles' proof of thm. 1.13; the reader is invited on a first reading to skip to sec. 3 to see how thm. 1.13 is deduced from the isomorphism criterion.

The proof we will give of Wiles' isomorphism criterion follows closely a presentation of Lenstra [Le].

2.1.2 Some examples

Before going further, it is good to pause and consider some examples of objects of $\mathcal{C}$ and the invariants associated to them. While logically independent of the proof, the examples should help the reader develop some intuition. It is therefore a good idea to work out these examples. (For an indication on how to compute the tangent spaces Φ_A, see the paragraph at the beginning of sec. 2.3.)

Example 1:

$$\begin{aligned} A &= \{(a,b) \in \mathbf{Z}_p \times \mathbf{Z}_p, a \equiv b \pmod{p^n}\} \\ &\simeq \mathbf{Z}_p[[T]]/(T(T-p^n)), \\ &\quad \pi(a,b) = a. \end{aligned}$$

$$\Phi_A \simeq \mathbf{Z}/p^n\mathbf{Z}, \qquad \eta_A = (p^n).$$

Example 2:

$$\begin{aligned} A &= \{(a,b,c) \in \mathbf{Z}_p \times \mathbf{Z}_p \times \mathbf{Z}_p, a \equiv b \equiv c \pmod{p}\} \\ &\simeq \mathbf{Z}_p[[X,Y]]/(X(X-p), Y(Y-p), XY), \\ &\quad \pi(a,b,c) = a. \end{aligned}$$

$$\Phi_A \simeq \mathbf{Z}/p\mathbf{Z} \times \mathbf{Z}/p\mathbf{Z}, \qquad \eta_A = (p).$$

Example 3:

$$\begin{aligned} A &= \mathbf{Z}_p[[X]]/(X^2), \\ &\quad \pi(f) = f(0). \end{aligned}$$

$$\Phi_A \simeq \mathbf{Z}_p, \qquad \eta_A = 0.$$

Example 4:

$$\begin{aligned} A &= \left\{ (a,b,c,d) \in \mathbf{Z}_p \times \cdots \times \mathbf{Z}_p, \begin{array}{c} a \equiv b \equiv c \equiv d \pmod{p}, \\ a+d \equiv b+c \pmod{p^2} \end{array} \right\} \\ &\simeq \mathbf{Z}_p[[X,Y]]/(X(X-p), Y(Y-p)), \\ &\quad \pi(a,b,c,d) = a. \end{aligned}$$

$$\Phi_A \simeq \mathbf{Z}/p\mathbf{Z} \times \mathbf{Z}/p\mathbf{Z}, \qquad \eta_A = (p^2).$$

Example 5:

$$\begin{aligned} A &= \mathbf{Z}_p[[X_1,\ldots,X_n]], \\ &\pi(f) = f(0). \end{aligned}$$

$$\Phi_A \simeq (\mathbf{Z}_p^n), \qquad \eta_A = (0).$$

Example 6: ($p \equiv -1 \pmod 4$).

$$\begin{aligned} A &= \{(a, b+ci) \in \mathbf{Z}_p \times \mathbf{Z}_p[i], a \equiv b \pmod{p^2}, c \equiv 0 \pmod p\} \\ &= \mathbf{Z}_p[[X]]/(X(X^2+p^2)). \\ &\pi(a,b+ci) = a. \end{aligned}$$

$$\Phi_A \simeq \mathbf{Z}/p^2\mathbf{Z}, \qquad \eta_A = (p^2).$$

Example 7:

$$\begin{aligned} A &= \mathbf{Z}_p[[T]]/(pT) = \mathbf{Z}_p \oplus \mathbf{F}_p T \oplus \mathbf{F}_p T^2 \oplus \cdots \\ &\pi(f) = f(0). \end{aligned}$$

$$\Phi_A \simeq (\mathbf{Z}/p\mathbf{Z}), \qquad \eta_A = (p).$$

2.2 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. The first two concern the behaviour of these invariants under morphisms (particularly surjective morphisms).

1. The assignment $A \mapsto \Phi_A$ is a functor from the category $\mathcal{C}$ to the category of finitely generated $\mathbf{Z}_p$-modules. Hence a morphism $A \longrightarrow B$ in $\mathcal{C}$ induces a homomorphism $\Phi_A \longrightarrow \Phi_B$ of $\mathbf{Z}_p$-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

$$\#\Phi_A \geq \#\Phi_B. \tag{4}$$

2. 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

$$\eta_A \subset \eta_B, \quad \text{i.e.,} \quad \#(\mathbf{Z}_p/\eta_A) \geq \#(\mathbf{Z}_p/\eta_B). \tag{5}$$

This is simply because in that case ϕ induces a map

$$\mathrm{Ann}\ker\pi_A \longrightarrow \mathrm{Ann}\ker\pi_B.$$

3. The third general property gives a relation between the invariants Φ_A and η_A. Observe that in all the examples of the previous section, we have $\#\Phi_A \in \eta_A$, whenever $\#\Phi_A$ is finite. In fact, this is true in general:

$$\#\Phi_A \geq \#(\mathbf{Z}_p/\eta_A). \tag{6}$$

The key behind proving this identity is to interpret $\#\Phi_A$ in terms of Fitting ideals. Namely, $\#\Phi_A$ (if it is finite) gives a generator for the Fitting ideal, $\mathrm{Fit}_{\mathbf{Z}_p}(\Phi_A)$ of Φ_A as a $\mathbf{Z}_p$ module. Hence

$$\begin{array}{rcl} \#\Phi_A & \in & \mathrm{Fit}_{\mathbf{Z}_p}(\Phi_A) \\ & = & \pi_A(\mathrm{Fit}_A(\ker\pi_A)) \\ & \subset & \pi_A(\mathrm{Ann}_A\ker\pi_A) = \eta_A. \end{array}$$

The first equality follows from the fact that, if M is any A module, then

$$\pi_A(\mathrm{Fit}_A(M)) = \mathrm{Fit}_{\mathbf{Z}_p}(M \otimes_A \mathbf{Z}_p),$$

where the tensor product is taken with respect to π_A. Applying this to $M = \ker\pi_A$, and observing that $\ker\pi_A \otimes_A \mathbf{Z}_p = (\ker\pi_A)/(\ker\pi_A)^2$, one finds the desired equality.

4. *Computing tangent spaces:*

Any object (A, π_A) in $\mathcal{C}$ can be expressed as a quotient of the object $U = \mathbf{Z}_p[[X_1, \ldots, X_n]]$ of example 5 with base point map given by $\pi_U(f) = f(0)$. For, 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 $\mathbf{Z}_p$-module of rank n which can be written down canonically as

$$\mathbf{Z}_pX_1 \oplus \mathbf{Z}_pX_2 \oplus \cdots \oplus \mathbf{Z}_pX_n,$$

the map $(\ker\pi_U)/(\ker\pi_U)^2$ 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

$$\Phi_A = \Phi_U/(\bar{f}_1,\ldots,\bar{f}_r).$$

This formula gives the most natural way of computing tangent spaces.

2.3 Complete intersections

We say that A is a *complete intersection* if it can be expressed as a quotient

$$A \simeq \mathbf{Z}_p[[X_1,\ldots,X_n]]/(f_1,\ldots,f_n)$$

where there are as many relations as there are variables.

The rings of examples 1, 3, 4, 6 and 7 are complete intersections (according to our definition), and the others are not. Observe that in the seven examples that are worked out, a ring A is a complete intersection precisely when $\#\Phi_A = \#(\mathbf{Z}_p/\eta_A)$.

An object A in $\mathcal{C}$ is said to be Gorenstein if

$$\mathrm{Hom}(A,\mathbf{Z}_p) \simeq A \quad \text{as } A\text{-modules.}$$

Proposition 2.2 *If A is a complete intersection, and A is a finitely generated $\mathbf{Z}_p$-module, then A is Gorenstein. (In particular, it is torsion-free as a $\mathbf{Z}_p$-module.)*

Remark: The condition that A be finitely generated is essential. For example, the ring A of example 7 is a complete intersection, but it is certainly not Gorenstein, since it is not even free as a $\mathbf{Z}_p$-module.

The remainder of this section will be devoted to proving prop. 2.2.

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 $(x_1,\ldots,x_n)$ generates a primary ideal of R, and $n = \dim R$, then $(x_1,\ldots,x_n)$ is called a *system of parameters* for R.

Lemma 2.3 *The sequence $(f_1,\ldots,f_n,p)$ is a system of parameters for $U = \mathbf{Z}_p[[X_1,\ldots,X_n]]$.*

Proof: The quotient ring $U/(f_1,\ldots,f_n,p)$ is local and is finitely generated as an $\mathbf{F}_p$-vector space; therefore every element in its maximal ideal is nilpotent.

A sequence $(x_1,\ldots,x_n)$ in a ring R is said to be a *regular sequence* if x_i is not a zero-divisor in $R/(x_1,\ldots,x_{i-1})$ for $i = 1,\ldots,n$.

Lemma 2.4 *The sequence $(f_1, \ldots, f_n)$ is a regular sequence for U.*

Proof: The ring U is Cohen Macaulay, since $p, X_1, \ldots, X_n$ is a system of parameters of U which is also a regular sequence. Hence, by theorem 17.4 (iii) of [Mat], $(f_1, \ldots, f_n, p)$ is a regular sequence in U. A fortiori, the sequence $(f_1, \ldots, f_n)$ is also a regular sequence.

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 $\underline{x} = (x_1, \ldots, x_n)$ of elements in its maximal ideal. This complex is defined to be the free graded differential algebra generated by symbols $u_1, \ldots, u_n$:

$$K_p(\underline{x}, R) := \oplus_{i_1 < i_2 < \cdots < i_p} R \cdot u_{i_1} \wedge \cdots \wedge u_{i_p},$$

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 2.5 *1. $H_0(\underline{x}, R) = R/(\underline{x})$.*

2. There is a long exact homology sequence

$$\cdots \longrightarrow H_p(\underline{x}, R) \longrightarrow H_p(\underline{x}, x_{n+1}, R) \longrightarrow H_{p-1}(\underline{x}, R) \overset{\pm x_{n+1}}{\longrightarrow}$$

$$H_{p-1}(\underline{x}, R) \longrightarrow H_{p-1}(\underline{x}, x_{n+1}, R) \longrightarrow H_{p-2}(\underline{x}, R) \longrightarrow \cdots$$

3. $H_p(\underline{x}, R)$ is annihilated by the ideal $(\underline{x})$, i.e., it has a natural $R/(\underline{x})$-module structure.

4. 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 2 and 3, see [Mat], th. 16.4. The assertion 4 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 4 follows. For more details on the Koszul complex and its relation to regular sequences, the reader can consult [Mat], especially §16.

Now, we turn to the proof of prop. 2.2, following a method of Tate which is explained in the appendix of [MRo]. For any ring R, write $R[[\underline{X}]] := R[[X_1, \dots, X_n]]$. Let $a_1, \dots, a_n$ be the images in A of $X_1, \dots, X_n$ by the natural map

$$\alpha : \mathbf{Z}_p[[\underline{X}]] \longrightarrow A = \mathbf{Z}_p[[\underline{X}]]/(f_1, \dots, f_n),$$

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:

$$(f_1, \dots, f_n) = (g_1, \dots, g_n)M,$$

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

$$\hom_{\mathbf{Z}_p}(A, \mathbf{Z}_p) \longrightarrow A.$$

We begin by constructing a surjective ($\mathbf{Z}_p$-linear) map

$$\hom_{\mathbf{Z}_p[[\underline{X}]]}(A[[\underline{X}]], \mathbf{Z}_p[[\underline{X}]]) \longrightarrow A.$$

Lemma 2.6 *(Tate): The function* $\Phi(f) = \alpha(f(D))$ *induces an isomorphism of* $\mathbf{Z}_p[[\underline{x}]]$*-modules*

$$\hom_{\mathbf{Z}_p[[\underline{X}]]}(A[[\underline{X}]], \mathbf{Z}_p[[\underline{X}]])/(g_1, \dots, g_n) \longrightarrow A.$$

Proof: By lemma 2.4, the sequence $(\underline{f}) = (f_1, \dots, f_n)$ is a regular $\mathbf{Z}_p[[\underline{X}]]$-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 prop. 2.5 that the Koszul complex $K(\underline{f})$ is a resolution of A by free $\mathbf{Z}_p[[\underline{X}]]$-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 $\mathbf{Z}_p[[\underline{X}]]$-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 $\mathbf{Z}_p[[\underline{x}]]$ into $A[[\underline{x}]]$, and letting

$$\Phi_1 : K_1(\underline{f}) \longrightarrow K_1(\underline{g})$$

be the map defined by

$$(\Phi_1(u_1), \dots, \Phi_1(u_n)) = (v_1, \dots, v_n)M,$$

and extending it by skew-linearity 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

$$\Phi_n(u_1 \wedge \dots \wedge u_n) = D \cdot v_1 \wedge \dots \wedge v_n.$$

Applying the functor $\hom_{\mathbf{Z}_p[[\underline{X}]]}(-, \mathbf{Z}_p[[\underline{X}]])$ to these two free resolutions, and taking the homology of the resulting complexes, we find that Φ induces an isomorphism on the cohomology, and in particular, on the nth cohomology:

$$\Phi_n : \hom_{\mathbf{Z}_p[[\underline{X}]]}(A[[\underline{X}]], \mathbf{Z}_p[[\underline{X}]])/(g_1, \dots, g_n) \xrightarrow{\simeq}$$
$$\hom(\mathbf{Z}_p[[\underline{X}]], \mathbf{Z}_p[[\underline{X}]])/(f_1, \dots, f_n) = A,$$

which is given explicitly by the formula:

$$\Phi_n(f) = \alpha(f(D)).$$

We finally come to the proof of prop. 2.2, which we can state in a more precise form.

Lemma 2.7 *The map $\Psi : \hom_{\mathbf{Z}_p}(A, \mathbf{Z}_p) \longrightarrow A$ defined by $\Psi(f) = \alpha(\tilde{f}(D))$, where $\tilde{f} : A[[\underline{X}]] \longrightarrow \mathbf{Z}_p[[\underline{X}]]$ 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

$$a = \alpha(a^{'}) \in A, \text{ with } a^{'} \in \mathbf{Z}_p[[\underline{X}]],$$

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 a $A[[\underline{X}]]$-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 $\mathbf{Z}_p[[\underline{X}]]$-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 $\mathbf{Z}_p$-basis of $\hom_{\mathbf{Z}_p}(A, \mathbf{Z}_p)$, then $\tilde{f}_1, \ldots, \tilde{f}_r$ is a $\mathbf{Z}_p[[\underline{X}]]$-basis for $\hom_{\mathbf{Z}_p[[\underline{X}]]}(A[[\underline{X}]], \mathbf{Z}_p[[\underline{X}]])$. 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 $\hom_{\mathbf{Z}_p}(A, \mathbf{Z}_p)$ and A are free $\mathbf{Z}_p$-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.

Examples:

1. Let $A \subset \mathbf{Z}_p \times \mathbf{Z}_p$ be the ring of example 1 in section 2.1.2. Then, $f = T^2 - p^n T$, and $g = T - (0, p^n)$. Hence, $f = (T - (p^n, 0))g$, so that $D = T - (p^n, 0)$. It follows that

$$\Phi(h) = \alpha(\tilde{h}(T - (p^n, 0))) = \alpha(Th(1,1) - h(p^n, 0)) = (0, p^n)h(1,1) - h(p^n, 0).$$

The reader can check that Φ is indeed an A-linear isomorphism from the A-module $\hom_{\mathbf{Z}_p}(A, \mathbf{Z}_p)$ to A.
2. Let $A = Z_p[\epsilon]/(\epsilon^2)$ be the ring of example 3 in section 2.1.2. Then, $f = T^2$, and $g = T - \epsilon$. Hence, $f = (T + \epsilon)g$, so that $D = T + \epsilon$. It follows that

$$\Phi(f) = \alpha(\tilde{f}(T + \epsilon)) = \alpha(Tf(1) + f(\epsilon)) = \epsilon f(1) + f(\epsilon).$$

A good way to gain insight into prop. 2.2 is to work out the isomorphism $\hom_{\mathbf{Z}_p}(A, \mathbf{Z}_p)$ for the complete intersection ring A of example 4 of sec. 2.1.2.

2.4 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 2.8 and 2.9 respectively.

Theorem 2.8 *Let $\phi : A \longrightarrow B$ be a surjective morphism of the category $\mathcal{C}$ with B a complete intersection. If ϕ induces an isomorphism $\Phi_A \longrightarrow \Phi_B$, and these modules are finite, then ϕ is an isomorphism.*

Remarks:
1. Let $A = \mathbf{Z}_p[[X, Y]]/(X(X-p), Y(Y-p))$ be the ring of example 4, let $B = \mathbf{Z}_p[[X, Y]]/(X(X-p), Y(Y-p), XY)$ 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 be a complete intersection is crucial for concluding that ϕ is an isomorphism.
2. The natural map

$$\mathbf{Z}_p[[X]]/(X^3) \longrightarrow \mathbf{Z}_p[[X]]/(X^2)$$

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: Let

$$\nu_B : \mathbf{Z}_p[[X_1, \ldots, X_n]] \longrightarrow B$$

be a surjective $\mathcal{C}$-morphism 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 they generate $\ker \pi_A$ as an A-module, by Nakayama's lemma. It follows (again by Nakayama) that the $\mathcal{C}$-morphism

$$\nu_A : \mathbf{Z}_p[[X_1, \ldots, X_n]] \longrightarrow A$$

defined by $\nu_A(X_i) = a_i$ is surjective. We claim that $\ker \nu_B \subset \ker \nu_A$ (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 $\mathbf{Z}_p$. 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 2.9 *Let $\phi : A \longrightarrow B$ be a surjective morphism of the category $\mathcal{C}$, with A being a complete intersection, and A and B finitely generated torsion free $\mathbf{Z}_p$-modules. If $\eta_A = \eta_B \neq 0$, then ϕ is an isomorphism.*

Remark: The torsion-freeness assumption on B is essential. For if A is a finitely generated complete intersection with $\eta_A \neq (0)$, and n is large enough, then $B = A/p^nA$ satisfies $\eta_A = \eta_B$, although the natural map $A \longrightarrow B$ is not surjective.

Proof: By proposition 2.2, A is Gorenstein, i.e., it satisfies

$$A^* = \hom_{\mathbf{Z}_p}(A, \mathbf{Z}_p) \simeq A \text{ as } A\text{-modules.}$$

Now, we observe that

$$\ker \pi_A \cap \mathrm{Ann}_A \ker \pi_A = 0, \tag{7}$$

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 $\mathbf{Z}_p$-torsion submodule of A, and therefore is 0.

It follows from (7) that 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 $\eta_A = \eta_B$, 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 (7) 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:

$$0 \longrightarrow \ker \phi \oplus \mathrm{Ann}_A \ker \pi_A \longrightarrow A. \tag{8}$$

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

$$B/(\mathrm{Ann}_B \ker \pi_B) \hookrightarrow \mathrm{end}_{\mathbf{Z}_p}(\ker \pi_B).$$

Hence, the exact sequence (8) splits over $\mathbf{Z}_p$. Taking $\mathbf{Z}_p$ duals in (8) and using the Gorenstein condition for A, we thus get an exact sequence of A-modules:

$$A \longrightarrow (\ker \phi)^* \oplus (\mathrm{Ann}_A \ker \pi_A)^* \longrightarrow 0.$$

Applying the functor $-\otimes_A \mathbf{F}_p$ (relative to the map $A \longrightarrow \mathbf{F}_p$), we find

$$1 = \dim_{\mathbf{F}_p}(A \otimes_A \mathbf{F}_p) \geq \dim_{\mathbf{F}_p}((\ker \phi)^* \otimes_A \mathbf{F}_p) + \dim_{\mathbf{F}_p}((\mathrm{Ann}_A \ker \pi_A)^* \otimes_A \mathbf{F}_p).$$

Since $\eta_A \neq 0$, it follows that $(\mathrm{Ann}_A \ker \pi_A)^* \otimes_A \mathbf{F}_p \neq 0$, and hence we must have

$$(\ker \phi)^* \otimes_A \mathbf{F}_p = 0.$$

Therefore by Nakayama's lemma and duality, $\ker \phi = 0$, which proves the theorem.

2.5 A resolution lemma

Although objects in $\mathcal{C}$ need not be complete intersections, they always can be "resolved" (in a weak sense) by a complete intersection, namely,

Theorem 2.10 *Let A be an object of $\mathcal{C}$. Then there is a morphism $\tilde{A} \longrightarrow A$ of $\mathcal{C}$ which induces an isomorphism $\Phi_{\tilde{A}} \longrightarrow \Phi_A$ and such that $\tilde{A}$ is a complete intersection. Moreover, if A is finitely generated over $\mathbf{Z}_p$, then $\tilde{A}$ can be chosen so as well.*

Proof: Write A as a quotient of $U = \mathbf{Z}_p[[X_1, \ldots, X_n]]$ (with $\pi_U : U \longrightarrow \mathbf{Z}_p$ the map which sends each X_i to 0.) Let $f_1, \ldots, f_n \in \ker \pi_U$ be such that $\bar{f}_1, \ldots, \bar{f}_n$ generate the kernel of $\Phi_U \longrightarrow \Phi_A$. Now letting

$$\tilde{A} = U/(f_1, \ldots, f_n)$$

gives the desired ring $\tilde{A}$.

It remains to show that, if A is a finitely generated $\mathbf{Z}_p$-module, then $\tilde{A}$ can be chosen to be finitely generated. This is not generally true of the ring $\tilde{A}$ constructed above, of course: we must take some care so that the n relations

$f_1, \ldots, f_n$ we leave in when defining $\tilde{A}$ are nice enough, so that $\tilde{A}$ is finitely generated.

Let $a_1, \ldots, a_n$ be $\mathbf{Z}_p$-module generators of the finitely generated module $\ker \pi_A$, and define a homomorphism ϕ from the polynomial ring

$$V = \mathbf{Z}_p[X_1, \ldots, X_n]$$

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 in the same way as before, 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

$$f_i + X_i^m h_i - X_i^{m+2},$$

and viewing these relations as belonging to the power series ring U instead of the polynomial ring V, we find that the ring

$$\tilde{A} = U/(f_1, \ldots, f_n)$$

has the desired properties:

1. 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.

2. The quotient $\tilde{A}$ is a finitely generated $\mathbf{Z}_p$-module, 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.

2.6 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 2.11 *Let A be an object of $\mathcal{C}$ which is a finitely generated torsion-free $\mathbf{Z}_p$-module. If $\#\Phi_A \leq \#(\mathbf{Z}_p/\eta_A) < \infty$, then A is a complete intersection.*

Proof: Let $\phi : \tilde{A} \longrightarrow A$ be the surjective morphism given by the resolution theorem (thm. 2.10). Then we have

$$\#(\mathbf{Z}_p/\eta_A) \geq (\#\Phi_A) = (\#\Phi_{\tilde{A}}) \geq \#(\mathbf{Z}_p/\eta_{\tilde{A}}),$$

where the first inequality is by assumption, the second by the choice of $\tilde{A}$, and the third is by the equation (6). On the other hand, by equation (5), we have

$$\#(\mathbf{Z}_p/\eta_{\tilde{A}}) \geq \#(\mathbf{Z}_p/\eta_A).$$

It follows that

$$\eta_A = \eta_{\tilde{A}},$$

so that ϕ is an isomorphism by thm. 2.9. It follows that A is a complete intersection.

2.7 Proof of Wiles' isomorphism criterion

Let us now recall the statement of Wiles' isomorphism criterion:

Theorem 2.12 *(Wiles) Let R and T be objects of $\mathcal{C}$ such that T is a finitely generated torsion-free $\mathbf{Z}_p$-module, and let $\phi : R \longrightarrow T$ be a surjective morphism. If*

$$\#\Phi_R \leq \#(\mathbf{Z}_p/\eta_T) < \infty,$$

then ϕ is an isomorphism.

Proof: We have:

$$\#(\mathbf{Z}_p/\eta_T) \leq \#\Phi_T \leq \#\Phi_R \leq \#(\mathbf{Z}_p/\eta_T),$$

where the first inequality is by equation (6), the second follows from the surjectivity of ϕ, and the third follows from the assumption of the theorem. Therefore,

$$\#\Phi_T = \#(\mathbf{Z}_p/\eta_T),$$

and hence T is a complete intersection. 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 thm. 2.8. This completes the proof.

2.8 The relative invariant $\eta_{T'/T}$

.

Let $\pi : T^{'} \longrightarrow T$ be a map of objects of $\mathcal{C}$. We will generalize slightly the invariant η of the previous sections, and introduce a relative invariant $\eta_{T'/T}$ defined by

$$\eta_{T'/T} = \pi(\mathrm{Ann}_{T'} \ker \pi),$$

which is an ideal of T.

We now describe a set-up in which the relative congruence ideal can be computed, in the case where T and $T^{'}$ are finite flat and reduced.

Suppose that $\Lambda^{'}$ is a free $T^{'}$-module and Λ is a free T-module of the same rank, k. The module Λ can also be viewed as a $T^{'}$-module via the map $T^{'} \longrightarrow T$. By choosing isomorphisms $\Lambda^{'} \simeq T^{'k}$ and $\Lambda \simeq T^k$, the map π induces a map $\alpha : \Lambda^{'} \longrightarrow \Lambda$ of $T^{'}$-modules.

Proposition 2.13 *If $\beta : \Lambda \longrightarrow \Lambda^{'}$ is an injective map of T'-modules such that $\Lambda^{'}/\beta\Lambda$ has no $\mathbf{Z}_p$-torsion, then*

$$\eta_{T'/T} = Ann_T(\Lambda/\alpha\beta\Lambda).$$

Proof: The map

$$\begin{array}{rcl} \hom_{T'}(T, T^{'}) & \longrightarrow & \mathrm{Ann}_{T'} \ker \pi \\ f & \mapsto & f(1) \end{array}$$

is an isomorphism, and hence $\hom_{T'}(\Lambda, \Lambda^{'}) \simeq M_n(\mathrm{Ann}_{T'} \ker \pi)$. It follows that after choosing an isomorphism of $\Lambda^{'}$ with $T^{'k}$, we have $\beta(\Lambda) \subset (\mathrm{Ann}_{T'} \ker \pi)^k$. Consider the exact sequence

$$0 \longrightarrow (\mathrm{Ann}_{T'} \ker \pi)^k/\beta\Lambda \longrightarrow \Lambda^{'}/\beta\Lambda \longrightarrow \Lambda^{'}/(\mathrm{Ann}_{T'} \ker \pi)^k \longrightarrow 0. \qquad (9)$$

Let $r(M) = \dim_{\mathbf{Q}_p}(M \otimes \mathbf{Q}_p)$ denote the $\mathbf{Z}_p$-rank of a finitely generated $\mathbf{Z}_p$-module M. Since $T^{'}$ is finite flat and reduced, we have

$$r(\mathrm{Ann}_{T'} \ker \pi) + r(\ker \pi) = r(T^{'}),$$

so that $r(\mathrm{Ann}_{T'} \ker \pi) = r(T)$. Since β is injective, it follows that

$$r((\mathrm{Ann}_{T'} \ker \pi)^k) = r(\Lambda) = r(\beta\Lambda),$$

and therefore the leftmost module in the exact sequence (9) is $\mathbf{Z}_p$-torsion. Therefore, it is trivial, since $\Lambda'/\beta\Lambda$ is assumed to be torsion-free. It follows that

$$\Lambda'/\beta\Lambda \simeq \Lambda'/(\mathrm{Ann}_{T'} \ker \pi)^k.$$

Applying α yields the proposition:

$$\Lambda/\alpha\beta\Lambda \simeq \Lambda/\eta_{T'/T}\Lambda.$$

2.9 Interpretation of $\eta_{T'/T}$ in the Gorenstein case

Assume now that the rings T' and T, in addition to being finite flat and reduced, are Gorenstein. Let $\pi^\vee : \underline{T}^\vee \longrightarrow T'^\vee$ be the dual map. By using the identifications of T and T' with their duals, we obtain an element

$$\pi \circ \pi^\vee : T \simeq T^\vee \longrightarrow T'^\vee \simeq T' \longrightarrow T,$$

which gives an element of T which is well-defined up to a unit.

Proposition 2.14 *The ideal of T generated by the image of $\pi \circ \pi^\vee$ is equal to $\eta_{T'/T}$.*

Proof: Let f be a T'-module generator of $T'^\vee$. The image of $T^\vee$ by $\pi^\vee$ is the set of all functions in $T'^\vee$ of the form $x \mapsto f(\lambda x)$, for some $\lambda \in \mathrm{Ann}(\ker \pi)$. The proposition follows.

This result allows us to give a relation between the relative congruence ideal and the absolute ones in the case where the rings T and T' are Gorenstein, namely:

Corollary 2.15 *If the rings T and T' are finite flat and reduced and satisfy the Gorenstein condition, then*

$$\eta_{T'} = \eta_T \cdot \pi_T(\eta_{T'/T}).$$

This formula allows us to compute the invariant $\eta_{T'}$ in terms of the relative invariant, and will be used later in studying the variation of η_T for Hecke rings when one increases the level.

3 Interpretation of Φ_{R_Σ} and η_{T_Σ}

To show theorem 1.13 Wiles must show that the inequality

$$\#\Phi_{R_\Sigma} \leq \#(\mathbf{Z}_5/\eta_{T_\Sigma}) \tag{10}$$

is satisfied for all finite sets Σ.

In section 3.1 we study the left hand side and show how to interpret Φ_{R_Σ} as the (dual of) a subgroup of a Galois cohomology group. We give an explicit formula for the right hand side in 3.2.

3.1 Interpretation of Φ_{R_Σ}

In this section we explain how one defines the symmetric square motive and the Selmer group associated to it, and how this Selmer group is (dual to) the tangent space Φ_{R_Σ}.

Let $X = X_0(17)$, and let $T = \mathrm{end}^0(T_5(X))$ denote the module of trace zero endomorphisms of the 5-adic Tate module $T_5(X)$; it is a free $\mathbf{Z}_5$-module of rank 3, equipped with a natural continuous $G_{\mathbf{Q}}$-action, defined by

$$\sigma(t) = \rho_0(\sigma)t\rho_0(\sigma)^{-1},$$

for all $t \in T$ and $\sigma \in G_{\mathbf{Q}}$. Define

$$V = T \otimes \mathbf{Q}_5, \quad A = T \otimes (\mathbf{Q}_5/\mathbf{Z}_5).$$

The vector space V and the $\mathbf{Z}_5$-divisible module A inherit a Galois action from T in the obvious way, and there is an exact sequence of $G_{\mathbf{Q}}$-modules

$$0 \longrightarrow T \longrightarrow V \longrightarrow A \longrightarrow 0.$$

Local behaviour at 5:
Since $X = X_0(17)$ is ordinary at 5, the module $T_5(X)$ is equipped with a two-step filtration

$$0 \subset \mathbf{Z}_5(1) \subset T_5(X)$$

preserved by the action of the inertia group I_5. Now, T contains a natural 1-dimensional I_5-stable subspace, consisting of the nilpotent endomorphisms which preserve this filtration. Call this subspace $T^o_{(5)}$. Likewise, define $V^o_{(5)}$ and $A^o_{(5)}$ by

$$V^o_{(5)} = T^o_{(5)} \otimes \mathbf{Q}_5, \qquad A^o_{(5)} = T^o_{(5)} \otimes \mathbf{Q}_5/\mathbf{Z}_5.$$

A more careful study shows that T is an ordinary representation, i.e., it is equipped with a 3-step filtration with 1-dimensional quotients on which I_5 acts by powers of the cyclotomic character. But we will not need this here.

Local behaviour at 17:
Likewise, the fact that $X_0(17)$ has split multiplicative reduction at 17 implies that $Ta_5(X)$ is equipped with a two-step filtration

$$0 \subset \mathbf{Z}_5(1) \subset Ta_5(X)$$

preserved by the action of the decomposition group D_{17}; this allows one to define as before a rank one D_{17}-stable submodule of T, consisting of the nilpotent endomorphisms preserving this filtration. Call this submodule $T^o_{(17)}$, and define as before $V^o_{(17)}$ and $A^o_{(17)}$ by

$$V^o_{(17)} = T^o_{(17)} \otimes \mathbf{Q}_5, \qquad A^o_{(17)} = T^o_{(17)} \otimes \mathbf{Q}_5/\mathbf{Z}_5.$$

The Selmer group $S_\Sigma(\mathbf{Q}, A)$:
Given a set of primes Σ not containing $\{5, 17\}$, we define a system J_r of subgroups of the local Galois cohomology groups

$$J_r \subset H^1(\mathbf{Q}_r, A),$$

for each prime r of $\mathbf{Q}$:
1. If $r \notin \Sigma \cup \{5, 17\}$, then

$$J_r = H^1(\mathbf{Q}_r^{nr}/\mathbf{Q}_r, A) := \ker\left(H^1(\mathbf{Q}_r, A) \longrightarrow H^1(I_r, A)\right).$$

2. If $r \in \Sigma$, then

$$J_r = H^1(\mathbf{Q}_r, A).$$

3. If $r = 17$, then

$$J_{17} = \ker\left(H^1(\mathbf{Q}_{17}, A) \longrightarrow H^1(\mathbf{Q}_{17}, A/A^o_{(17)})\right).$$

4. If $r = 5$, then

$$J_5 = \ker\left(H^1(\mathbf{Q}_5, A) \longrightarrow H^1(I_5, A/A^o_{(5)})\right).$$

Define the Selmer group (relative to Σ) by the formula

$$S_\Sigma(\mathbf{Q}, A) = \ker\left(H^1(\mathbf{Q}, A) \longrightarrow \prod_r H^1(\mathbf{Q}_r, A)/J_r\right).$$

One can also define the Selmer groups $S_\Sigma(\mathbf{Q}, A_{5^n}) \subset H^1(\mathbf{Q}, A_{5^n})$, in the obvious way. It follows from the fact that $H^0(\mathbf{Q}, A_5) = 0$ that $S_\Sigma(\mathbf{Q}, A_5) = S_\Sigma(\mathbf{Q}, A)_5$, and that

$$S_\Sigma(\mathbf{Q}, A) = \varinjlim S_\Sigma(\mathbf{Q}, A_{5^n}).$$

The universal deformation class:
Recall that R_Σ is the universal deformation ring associated to the set of primes Σ, and let $\mathcal{P}_{R_\Sigma}$ be the kernel of the base point map $R_\Sigma \longrightarrow \mathbf{Z}_5$.

There is a natural *split* exact sequence

$$1 \longrightarrow M_2(\Phi_{R_\Sigma}) \longrightarrow \mathbf{GL}_2(R_\Sigma/\mathcal{P}^2_{R_\Sigma}) \longrightarrow \mathbf{GL}_2(\mathbf{Z}_5) \longrightarrow 1,$$

where the action of $\mathbf{GL}_2(\mathbf{Z}_5)$ on the space of matrices $M_2(\Phi_{R_\Sigma})$ is by conjugation. By the standard cohomological construction, the Galois representation

$$\rho' : G_\mathbf{Q} \longrightarrow \mathbf{GL}_2(R_\Sigma/\mathcal{P}^2_{R_\Sigma})$$

coming from the universal deformation gives rise to a "universal deformation class"

$$u_\Sigma \in H^1(\mathbf{Q}, M_2(\Phi_{R_\Sigma})).$$

The class u_Σ is constructed explicitly as follows. Let $\tilde{\rho}_0$ be the natural lift of ρ_0 to $\mathbf{GL}_2(R_\Sigma/\mathcal{P}^2_{R_\Sigma})$. Then

$$u_\Sigma(\sigma) = \tilde{\rho}_0(\sigma)\rho'(\sigma)^{-1}.$$

Since $\det(u_\Sigma(\sigma)) = 1$, it follows that $u_\Sigma(\sigma)$ can be viewed as belonging to the Lie algebra $sl_2(\Phi_{R_\Sigma}) = T \otimes \Phi_{R_\Sigma}$. The class u_Σ sets up a homomorphism

$$\phi_\Sigma : \mathrm{hom}(\Phi_{R_\Sigma}, \mathbf{Q}_5/\mathbf{Z}_5) \longrightarrow H^1(\mathbf{Q}, A)$$

through the natural rule

$$\phi_\Sigma(f) = f(u_\Sigma).$$

Claim 3.1 *The homomorphism ϕ_Σ gives an isomorphism*

$$\phi_\Sigma : \hom(\Phi_{R_\Sigma}, \mathbf{Q}_5/\mathbf{Z}_5) \xrightarrow{\simeq} S_\Sigma(\mathbf{Q}, A).$$

Proof: First, one must show that ϕ_Σ actually maps to the Selmer group. This follows from the definitions. At the prime 5, the assumption that the deformation $\rho^{'}$ is admissible means that

$$\rho^{'}|_{I_5} = \begin{pmatrix} \epsilon & \Psi^{'} \\ 0 & 1 \end{pmatrix}, \quad \tilde{\rho}_o|_{I_5} = \begin{pmatrix} \epsilon & \tilde{\Psi}_o \\ 0 & 1 \end{pmatrix},$$

and hence the restriction of the universal class u_Σ to I_5 belongs to

$$H^1(I_5, T^o_{(5)} \otimes \Phi_{R_\Sigma}).$$

Likewise, one checks that the restriction of the universal class u_Σ to the decomposition group D_{17} belongs to $H^1(\mathbf{Q}_{17}, T^o_{(17)} \otimes \Phi_{R_\Sigma})$, so that $\Phi_\Sigma(f)$ does belong to $S_\Sigma(A)$. The fact that ϕ_Σ is an isomorphism is a formal consequence of the universality of R_Σ. For, distinct elements of $\hom(\Phi_{R_\Sigma}, \mathbf{Q}_5/\mathbf{Z}_5)$ giving rise to the same element in $S_\Sigma(\mathbf{Q}, A)$ would give distinct maps $R_\Sigma \longrightarrow R$ (for some local ring R) corresponding to equivalent (i.e., equal) deformations, contradicting the uniqueness clause in the definition of the universal deformation ring. This proves injectivity. Surjectivity is proved in exactly the same way.

3.2 A formula for η_{T_Σ}

Recall that

$$f(q) = \sum_n a_n q^n$$

is the normalized eigenform with integer coefficients corresponding to $X = X_0(17)$. The following gives a precise formula for $\#(\mathbf{Z}_5/\eta_{T_\Sigma})$.

Theorem 3.2 *Up to units in $\mathbf{Z}_5^*$, we have:*

$$\#(\mathbf{Z}_5/\eta_{T_\Sigma}) = \prod_{q\in\Sigma} (q-1)(a_q^2 - (q+1)^2).$$

Observe that when $\Sigma = \emptyset$, then $\mathbf{T}_\emptyset$ is the $\mathbf{Z}_5$-algebra of Hecke operators acting on $S_2(17, \mathbf{Z}_5)$, and hence, since $X_0(17)$ is of genus 1,

$$\mathbf{T}_\emptyset \simeq \mathbf{Z}_5.$$

The base point map $\mathbf{T}_\emptyset \longrightarrow \mathbf{Z}_5$ is the identity map; hence, in that case

$$\eta_{T_\emptyset} = (1).$$

We prove theorem 3.2 by induction on the size of Σ. Assume the theorem is true for Σ, and let $\Sigma^{'} = \Sigma \cup \{q\}$, where q is a prime not in Σ. We view these rings, equipped with the base point maps arising from the forms f_Σ and $f_{\Sigma'}$, as objects in the category $\mathcal{C}$ introduced in section 2. To apply induction, one needs a surjective map $\mathbf{T}_{\Sigma'} \longrightarrow \mathbf{T}_\Sigma$ between these objects. We being by constructing such a map.

3.2.1 A map $\mathbf{T}_{\Sigma'} \longrightarrow \mathbf{T}_\Sigma$

Since $N_{\Sigma'} = q^2 N_\Sigma$, there is a natural "degeneracy map"

$$\xi : X_0(N_{\Sigma'}) \longrightarrow X_0(N_\Sigma)^3,$$

given by the formula:

$$\xi(\tau) = (\tau, q\tau, q^2\tau).$$

This map induces by Pic functoriality a map i:

$$i : J_0(N_\Sigma)^3 \longrightarrow J_0(N_{\Sigma'}).$$

Define a map β_0:

$$\beta_0 : J_0(N_\Sigma) \longrightarrow J_0(N_\Sigma)^3$$

by the formula

$$\beta_0(P) = (qP, -T_q P, P).$$

The map β_0 is a map of $\mathbf{T}(\Sigma)$-modules, and i commutes with the obvious action of $\mathbf{T}^0(\Sigma^{'})$ - in fact i respects the natural action of all the Hecke operators, except U_q which does not act on $J_0(N_\Sigma)^3$ in any obvious way. However, we do have:

Lemma 3.3 *The image of $J_0(N_\Sigma)^3$ in $J_0(N_{\Sigma'})$ under i is stable under the action of the Hecke operator U_q, and U_q induces the endomorphism*

$$\begin{pmatrix} T_q & q & 0 \\ -1 & 0 & q \\ 0 & 0 & 0 \end{pmatrix}$$

acting on $J_0(N_\Sigma)^3$.

Proof: This is verified by a direct calculation. It is convenient to view elements of $J_0(N_\Sigma)$ as formal sums of degree 0 of objects of the form $[C]$, where C is an elliptic curve equipped with the appropriate N_Σ level structure, and elements of $J_0(N_{\Sigma'})$ as sums of objects of the form $[C_1 \to C_2 \to C_3]$, where the C_i are curves with level N_Σ structure and the arrows are cyclic q-isogenies whose composite is cyclic of degree q^2. The formula for i on such objects is:

$$i([C_1],[C_2],[C_3]) = \sum_{A,B}[C_1 \to A \to B] + \sum_{A,B}[A \to C_2 \to B] + \\ + \sum_{A,B}[A \to B \to C_3], \tag{11}$$

where the sums in each case are taken over all possible A, B so that the composite arrow is cyclic of degree q^2. (Hence, each sum contains $q(q+1)$ distinct terms.) The formula for U_q on $J_0(N_{\Sigma'})$ is

$$U_q([A \to B \to C]) = \sum_D [B \to C \to D], \tag{12}$$

where again the sum is taken over the q possible D such that the composite isogeny is cyclic of degree q^2. Using equations (11) and (12), one checks directly that $i(J_0(N_\Sigma)^3)$ is stable under U_q, and in fact that

$$U_q(i([C_1],[C_2],[C_3])) = i(T_q[C_1] + q[C_2], -[C_1] + q[C_3], 0),$$

so that U_q acts by the matrix

$$\begin{pmatrix} T_q & q & 0 \\ -1 & 0 & q \\ 0 & 0 & 0 \end{pmatrix},$$

as was to be shown.

Let now $\beta = i\beta_0$. From lemma 3.3, one sees that image of β is stable under U_q, and in fact, is killed by U_q. Hence, one can define a natural ring homomorphism

$$\pi : \mathbf{T}(\Sigma^{'}) \longrightarrow \mathbf{T}(\Sigma)$$

sending an operator to its restriction to $\beta(J_0(N_\Sigma))$.

Lemma 3.4 *The map $\pi : \mathbf{T}(\Sigma^{'}) \otimes \mathbf{Z}_5 \longrightarrow \mathbf{T}(\Sigma) \otimes \mathbf{Z}_5$ is surjective, and maps the ideal $\mathcal{M}_{\Sigma'}$ to the ideal $\mathcal{M}_\Sigma$.*

Proof: To show surjectivity, since $\mathbf{T}(\Sigma) \otimes \mathbf{Z}_5$ is generated by the Hecke operators T_l and U_l for $l \in \{17\} \cup \Sigma$, it suffices to show that these operators are in the image of π. The operators T_l with $l \neq q$ and U_l are just the images of the corresponding operators in $\mathbf{T}(\Sigma^{'}) \otimes \mathbf{Z}_5$. As for the operator T_q, it follows directly from the Chebotarev density theorem (using the fact that $\mathbf{T}(\Sigma) \otimes \mathbf{Z}_5$ is 5-adically topologically complete) that it is in the ring generated by the good Hecke operators T_l, $l \neq q$. Finally, π maps $\mathcal{M}_\Sigma$ to $\mathcal{M}_{\Sigma'}$, since it sends $T_l - a_l$ to $T_l - a_l$ ($l \notin \{17\} \cup \Sigma$) and sends U_l to U_l if $l \neq q$, and to 0 otherwise.

Hence, by completing at the ideals $\mathcal{M}_{\Sigma'}$ and $\mathcal{M}_\Sigma$, we get a surjective map (which by abuse of notation we also call π):

$$\pi : \mathbf{T}_{\Sigma'} \longrightarrow \mathbf{T}_\Sigma,$$

and we can define a relative congruence ideal associated to this map, as in sec. 2.8.

3.2.2 Computing $\eta_{\mathbf{T}'/\mathbf{T}}$

Let $\mathbf{T} = \mathbf{T}_\Sigma$ and $\mathbf{T}^{'} = \mathbf{T}_{\Sigma'}$ be the Hecke rings at level Σ and $\Sigma^{'}$. By applying the induction hypothesis and cor. 2.15 of sec. 2.8, we are reduced to showing that

$$\pi_{\mathbf{T}}(\eta_{\mathbf{T}'/\mathbf{T}}) = ((q-1)(a_q^2 - (q+1)^2).$$

In fact, we will show:

$$\eta_{\mathbf{T}'/\mathbf{T}} = ((q-1)(T_q^2 - (q+1)^2)),$$

where $T_q \in \mathbf{T}$ is the qth Hecke operator.

Let the maps i, β_0, and β be as above, and let j, α_0 and α denote their duals.

The maps i, j, α_0 and β_0 induce maps on the 5-adic Tate modules of the associated abelian varieties, which, by abuse of notation, we will denote by the same letters, and we have a sequence of maps:

$$\begin{tikzcd}
Ta_5(J_0(N_\Sigma))^3 \arrow[r, "i"] & Ta_5(J_0(N_{\Sigma'})) \arrow[r, "j"] & Ta_5(J_0(N_\Sigma))^3 \arrow[d, "\alpha_0"] \\
Ta_5(J_0(N_\Sigma)) \arrow[u, "\beta_0"] & & Ta_5(J_0(N_\Sigma))
\end{tikzcd} .$$

Now let

$$\Lambda := Ta_5(J_0(N_\Sigma)) \otimes_{\mathbf{T}(\Sigma)} \mathbf{T}_\Sigma, \quad \Lambda^{'} := Ta_5(J_0(N_{\Sigma'})) \otimes_{\mathbf{T}(\Sigma')} \mathbf{T}_{\Sigma'}.$$

The maps α and β induce maps on these tensored modules, which, by abuse of notation, we will again call α and β. Hence, we have a diagram of morphisms:

$$\Lambda \xrightarrow{\beta} \Lambda^{'} \xrightarrow{\alpha} \Lambda.$$

Note that β and α are maps of $\mathbf{T}_{\Sigma'}$-modules when we endow Λ with its $\mathbf{T}_{\Sigma'}$-module structure coming from the map $\mathbf{T}_{\Sigma'} \longrightarrow \mathbf{T}_\Sigma$. Furthermore, we have the following key propositions:

Theorem 3.5 *The rings $\mathbf{T}_\Sigma$ and $\mathbf{T}_{\Sigma'}$ are Gorenstein, and $\Lambda \simeq \mathbf{T}_\Sigma^2$ and $\Lambda^{'} \simeq \mathbf{T}_{\Sigma'}^2$.*

This theorem was proved first by Mazur in [Mz1] in the case of $J_0(N)$, N prime, and has since been extended by a number of people, including Wiles in [Wi]; cf. also [Ed].

Theorem 3.6 *The module $\Lambda^{'}/\beta\Lambda$ has no $\mathbf{Z}_5$-torsion.*

Since $Ta_5(J_0(N_\Sigma)^3)/\beta_0(Ta_5(J_0(N_\Sigma)))$ is torsion free, it suffices to show that $Ta_5(J_0(N_{\Sigma'})/i(Ta_5(J_0(N_\Sigma)^3)$ is torsion free, at least after tensoring with $\mathbf{T}_{\Sigma'}$. This was proved by Ribet [Ri1] in the case of the map

$$i : J_0(N)^2 \longrightarrow J_0(Nq), \quad \gcd(q, N) = 1$$

induced by the obvious two degeneracy maps, using some results of Ihara. Building on Ribet's result, Wiles has extended it to cover the case he needs; cf. [Wi].

Thanks to thm. 3.5 and thm. 3.6, we are in the situation of prop. 2.13, so that it suffices now to compute the modules $\Lambda/\alpha\beta\Lambda$.

Lemma 3.7 *The map $j \circ i$, viewed as an endomorphism of $Ta_5(J_0(N_\Sigma))^3 \otimes \mathbf{T}_\Sigma$ in $M_3(\mathbf{T}_\Sigma)$, is equal to*

$$\left(\begin{array}{ccc} q(q+1) & qT_q & T_q^2-(q+1) \\ qT_q & q(q+1) & qT_q \\ T_q^2-(q+1) & qT_q & q(q+1) \end{array} \right).$$

Proof: One proceeds as in the proof of lemma 3.3, using the explicit formula for i given there, combined with the formula

$$j([A \to B \to C]) = ([A],[B],[C]).$$

Finally we come to:

Proposition 3.8 $\alpha\beta = -q(q-1)(T_q^2-(q+1)^2)$.

Proof: One notes that α_0, the dual of β_0, is given by the formula

$$\alpha_0([C_1],[C_2],[C_3]) = q[C_1] - T_q[C_2] + [C_3].$$

$$\begin{aligned} \alpha\beta([C]) &= \alpha_0 ji\beta_0([C]) = \alpha_0 ji(q[C], -T_q[C], [C]) \\ &= \alpha_0(-(q-1)(T_q^2-(q+1)^2)[C], 0, 0) \\ &= -q(q-1)(T_q^2-(q+1)^2)[C]. \end{aligned}$$

Corollary 3.9 $\eta_{\mathbf{T}'/\mathbf{T}} = ((q-1)(T_q^2-(q+1)^2))$.

Proof: By prop. 3.8, we have

$$\mathrm{Ann}_{\mathbf{T}}(\Lambda/\alpha\beta\Lambda) = ((q-1)(T_q^2-(q+1)^2)),$$

since $q \neq 5$ is a unit in $\mathbf{T}_\Sigma$. The proof follows from prop. 2.13.

The proof of thm. 3.2 now follows from cor. 3.9 and cor. 2.15.

3.3 Relation with the Bloch-Kato conjectures

In this section we mention briefly the relation between η_{T_Σ} and special values of L-functions, and the relation between Wiles' inequality and the Bloch-Kato conjecture. This section is not logically needed for the proof of thm. 1.13.

Let us drop for this section our running assumption that X is the modular elliptic curve $X_0(17)$. Let X be an arbitrary (semistable, to simplify) modular elliptic curve, and define algebraic numbers α_q and β_q for each prime q by the conditions

$$\alpha_q + \beta_q = a_q, \qquad \alpha_q\beta_q = q, \quad \text{if } X \text{ has good reduction at } q,$$

$$\alpha_q = \pm 1, \qquad \beta_q = 0 \text{ otherwise},$$

where α_q is 1 precisely when X has split multplicative reduction at q. Now define the local L-function associated to T by

$$L_q(T,s) = (1-\alpha_q^2 q^{-s})^{-1}(1-\beta_q^2 q^{-s})^{-1}(1-\alpha_q\beta_q q^{-s})^{-1}.$$

Finally, we define the global L-function associated to T and the set Σ by

$$L_\Sigma(T,s) = \prod_{q\notin\Sigma} L_q(T,s).$$

This function extends to an entire function on the complex plane, and satisfies a functional equation interchanging s and $3-s$.

By using Rankin's method and some techniques of Shimura, Sturm [St] has shown that $L_\Sigma(T,2)$ is a rational multiple of a transcendental period Ω,

$$\Omega = \int_{X(\mathbf{C})} \omega \wedge \bar{\omega},$$

where ω is a Néron differential attached to X.

If $x \in \mathbf{Q}$, let $[x]_5 = 5^{ord_5(x)}$ denote the 5-part of x. Hida, using the fundamental work of Ribet and others, has succeeded in relating the special value $L_\Sigma(T,2)$ to congruences between modular forms in a very precise way (usually in the case where Σ is minimal). His ideas, combined with the calculations of sec. 3.2, lead to the following relation:

Theorem 3.10

$$\#(\mathbf{Z}_5/\eta_{T_\Sigma}) = \left[\frac{L_\Sigma(T,2)}{\Omega}\right].$$

(Of course, one has a similar formula with any prime p replacing 5.) Thanks to thm. 3.10, Wiles conjectured inequality (10) can be reformulated as:

$$\#S_\Sigma(\mathbf{Q}, A) \leq \left[\frac{L_\Sigma(T,2)}{\Omega}\right].$$

This fits into the framework of the general conjectures of Bloch and Kato on the special values of the L-functions attached to motives which are a vast generalization of Dirichlet's class number formula and the Birch Swinnerton Dyer conjecture.

Such a reformulation is striking, since it gives a justification for the Shimura-Taniyama conjecture from a rather unexpected point of view. On the other hand, a proof of the inequality could be a priori quite difficult - one just has to think of how elusive the Birch Swinnerton-Dyer conjecture remains, even for modular elliptic curves!

Fortunately, the Bloch-Kato conjecture for the symmetric square is "easier" than the Birch Swinnerton-Dyer conjecture, in one important respect: the special value of the symmetric square L-function is always non-zero at the point of interest. Hence the difficulties which are caused by the unpredictable behaviour of the order of vanishing at $s = 1$ for the L-function of an elliptic curve, do not arise in this situation.

Furthermore, the ideas of Wiles and Flach allow us to come to grips with the Selmer groups of the symmetric square in many cases, (and, for example, to prove their finiteness). Naturally, one hopes that these techniques can be pushed further to prove the analogue of the class number formula in this setting, and hence, the full Shimura-Taniyama conjecture.

4 Proof of the inequality $\#\Phi_{R_\Sigma} \leq \#(\mathbf{Z}_5/\eta_{T_\Sigma})$

4.1 Reduction to the case $\Sigma = \emptyset$

Wiles first shows that the inequality

$$\#(\Phi_{R_\Sigma}) \leq \#(\mathbf{Z}_5/\eta_{T_\Sigma})$$

follows from the corresponding inequality for the minimal set $\Sigma = \emptyset$, by a simple induction argument.

Assume the inequality is true for Σ, and let $\Sigma^{'} = \Sigma \cup \{q\}$ for some prime number q. The exact sequence

$$0 \longrightarrow S_{\Sigma}(\mathbf{Q}, A) \longrightarrow S_{\Sigma'}(\mathbf{Q}, A) \longrightarrow H^1(I_q, A)^{D_q}$$

implies that

$$\#\Phi_{\Sigma'} \leq (\#\Phi_{\Sigma})(\#H^1(I_q, A)^{D_q}), \tag{13}$$

and our explicit formula for $\eta_{T_{\Sigma}}$ of thm. 3.2 implies that

$$\#(\mathbf{Z}_5/\eta_{T_{\Sigma'}}) = \#(\mathbf{Z}_5/\eta_{T_{\Sigma}})((q+1)^2 - a_q^2)(q-1). \tag{14}$$

Combining equations (13) and (14), we are reduced to showing that

$$\#H^1(I_q, A)^{D_q} \leq [((q+1)^2 - a_q^2)(q-1)].$$

This follows from an explicit calculation; since I_q acts trivially on A,

$$H^1(I_q, A)^{D_q} = \hom_{D_q}(I_q, A) = \hom_{D_q}(\mathbf{Z}_5(1), A).$$

Since I_q acts trivially on $\mathbf{Z}_5(1)$ and A, the group D_q acts on both via D_q/I_q, which is topologically generated by the Frobenius element Frob_q. The eigenvalue of Frob_q on $\mathbf{Z}_5(1)$ is q, and on A the eigenvalues are

$$1, \quad \alpha_q^2/q, \quad \beta_q^2/q.$$

Hence, we find

$$\#H^1(I_q, A)^{D_q} = [(q-1)(q^2 - \alpha_q^2)(q^2 - \beta_q^2)]. \tag{15}$$

Using the identity $\alpha_q^2 + \beta_q^2 = a_{q^2} = a_q^2 - 2q$, we find that the right hand side is $[(q-1)((q+1)^2 - a_q^2)]$, as was to be showed.

Remark: In the end, all the inequalities that Wiles proves turn out to be actual equalities. Hence, the above calculation indicates that the reduction map

$$S_{\Sigma'}(\mathbf{Q}, A) \longrightarrow H^1(I_q, A)^{D_q}$$

is *surjective* for all choices of $\Sigma^{'}$. This assertion, in the case $\Sigma^{'} = \{q\}$ and $\Sigma = \emptyset$ for certain primes q, turns out to be a key ingredient in bounding the order of $S_{\emptyset}(\mathbf{Q}, A)$, as we will explain in the next section.

4.2 Proof of the inequality for $\Sigma = \emptyset$

We are reduced to showing the inequality (10) for the minimal set $\Sigma = \emptyset$.

Theorem 4.1 $\#S_{\emptyset}(\mathbf{Q}, A) = 1$.

There are at least three ways in which this can be proved:

1. The ad hoc method: Taking cohomology of the exact sequence

$$0 \longrightarrow A_5 \longrightarrow A \longrightarrow A \longrightarrow 0,$$

and using the fact that $H^0(\mathbf{Q}, A) = 0$, gives the isomorphism:

$$H^1(\mathbf{Q}, A)_5 \simeq H^1(\mathbf{Q}, A_5).$$

Let $K/\mathbf{Q}$ be the splitting field of A_5, and let $G = \mathrm{Gal}(K/\mathbf{Q}) \simeq \mathbf{GL}_2(\mathbf{F}_5)$ be its Galois group. As we will see later (lemma 4.9) the cohomology group $H^1(G, A_5)$ vanishes, and hence the inflation-restriction sequence gives an injection

$$H^1(\mathbf{Q}, A_5) \longrightarrow \hom_G(\mathrm{Gal}(K_{(5)}/K), A_5),$$

where $K_{(5)}$ is the maximal abelian extension of K of exponent 5, and the action of G is the natural one. Let $K'_{(5)}$ be the abelian extension of exponent 5 which is unramified outside of 5 and 17. Then $S_{\emptyset}(\mathbf{Q}, A_5)$ injects into

$$\hom_G(\mathrm{Gal}(K'_{(5)}/K), A_5),$$

and one can therefore hope to control $S_{\emptyset}(\mathbf{Q}, A_5)$ by controlling the size of this group, which is (dual to) a piece of the 5-part of a well-defined class group of K. Since the field K is of degree 480 over $\mathbf{Q}$, this calculation seems rather daunting. It would be interesting to see if it can be carried out in practice, (which would probably require a more careful description of the image of $S_{\emptyset}(\mathbf{Q}, A)$ as well as some theoretical insights into how the class group of the field K behaves.) At any rate, it is worth noting that at this stage we have reduced the proof of thm 1.1 to a finite amount of (machine) calculation.

2. Flach's approach: In [Fl], Flach shows that if the module A arises from the symmetric square of any (semi-stable) modular elliptic curve X, then $\deg\phi$ annihilates $\#S_{\Sigma}(\mathbf{Q}, A)$, where ϕ is a minimal degree of a modular parametrization from $X_0(N)$ to X. In our case, this modular parametrization is the identity map $X_0(17) \longrightarrow X_0(17)$ of degree 1. Hence $S_\Sigma(\mathbf{Q}, A)$ is trivial.

Flach's argument proceeds by constructing explicit "Kolyvagin-type" cohomology classes

$$c(l) \in H^1(\mathbf{Q}, T^*),$$

where $T^* = \hom(T, \mathbf{Z}_5(1))$ is the Kummer dual of T. Roughly speaking, these classes are constructed as follows: the degeneracy maps of section 3.2 give a map of $X_0(17l)$ into $X_0(17) \times X_0(17)$ whose image is the Hecke correspondance T_l on the surface $X_0(17) \times X_0(17)$. Flach defines a certain modular unit u on $X_0(17l)$ whose divisor becomes trivial on the image T_l. The divisor T_l, together with the unit u, gives rise to a class in a certain algebraic K-group $H^1(X_0(17) \times X_0(17), \mathcal{K}_2)$, which maps to the étale cohomology group

$$H^3_{et}(X_0(17) \times X_0(17), \mathbf{Z}_5(2)).$$

The Hochshild-Serre spectral sequence, combined with the fact that

$$H^0(\mathbf{Q}, H^3_{et}(X_0(17) \times X_0(17), \mathbf{Z}_5(2))) = 0$$

allows Flach to transfer this class to the group

$$H^1(\mathbf{Q}, H^2_{et}(X_0(17) \times X_0(17), \mathbf{Z}_5(2))),$$

which maps, via the Künneth projections, to

$$H^1(\mathbf{Q}, H^1_{et}(X_0(17), \mathbf{Z}_5(1))^{\otimes 2}).$$

By projecting to the symmetric tensors, Flach obtains his class

$$c(l) \in H^1(\mathbf{Q}, T^*).$$

By a careful and delicate analysis, Flach shows that $c(l)$ is ordinary at 5 and ramified only at l, and that its restriction to the inertia group I_l, on the other hand, generates $H^1(I_l, T^*)^{D_l}$, at least for sufficiently many l, (those for which 5 does not divide $((l+1)^2 - a_l^2)$). An application of the local Tate duality and the global reciprocity law of class field theory shows that for all these l, and for all $s \in S_\Sigma(\mathbf{Q}, A)$, the restriction of s to the decomposition group D_l is trivial. Such stringent local conditions force the Selmer group

$S_\Sigma(\mathbf{Q}, A)$ to be trivial, by an application of the Chebotarev density theorem. All of this is very well explained in Flach's paper [Fl].

Flach's method is inspired by the work of Kolyvagin, who constructed ramified classes $c(l)$ in different contexts to obtain annihilators of the appropriate Selmer groups [Ko1], thereby establishing some striking cases of the Birch Swinnerton Dyer conjecture. In his seminal paper [Ko] on Euler systems, Kolyvagin showed how to strengthen this argument to obtain actual upper bounds on the Selmer groups, by constructing systems of cohomology classes $c(l_1, \ldots, l_k)$, extending his classes $c(l)$, satisfying a subtle compatibility property relating the restriction to D_{l_k} of $c(l_1, \ldots, l_k)$ to that of the previous class $c(l_1, \ldots, l_{k-1})$.

It is a tantalizing question to see if Flach's cohomology classes $c(l)$ can be likewise extended to an Euler system, and whether this approach can yield a proof of the inequality (10) in more general contexts.

3. Wiles' approach: Wiles proves the inequality (10) for $X_0(17)$ by a completely different route, which relies crucially on the fact that the Hecke ring $\mathbf{T}_\emptyset$ is a local complete intersection.

We will explain Wiles' proof here in some detail.

4.2.1 The group $S_\Sigma^{mod}(\mathbf{Q}, A)$

For any finite set Σ not containing 5 and 17, let

$$S_\Sigma^{mod}(\mathbf{Q}, A) := \Phi_{\mathbf{T}_\Sigma}^\vee,$$

where $M^\vee = \hom(M, \mathbf{Q}_5/\mathbf{Z}_5)$ denotes the Pontrjagin dual of a $\mathbf{Z}_5$-torsion module. The surjective map $\Phi_{R_\Sigma} \longrightarrow \Phi_{\mathbf{T}_\Sigma}$ gives, upon passing to the duals, an inclusion

$$S_\Sigma^{mod}(\mathbf{Q}, A) \subset S_\Sigma(\mathbf{Q}, A).$$

The group $S_\Sigma^{mod}(\mathbf{Q}, A)$ should be thought of as the part of $S_\Sigma(\mathbf{Q}, A)$ which "comes from modular forms".

If $\Sigma_1 \subset \Sigma_2$ are finite sets of primes not containing 5 and 17, then, by definition there is an exact sequence

$$0 \longrightarrow S_{\Sigma_1}(\mathbf{Q}, A) \longrightarrow S_{\Sigma_2}(\mathbf{Q}, A) \longrightarrow \oplus_{\Sigma_2 - \Sigma_1} H^1(I_q, A)^{D_q}.$$

One could expect a similar statement for the modular Selmer groups, i.e., that the sequence

$$0 \longrightarrow S^{mod}_{\Sigma_1}(\mathbf{Q}, A) \longrightarrow S^{mod}_{\Sigma_2}(\mathbf{Q}, A) \longrightarrow \oplus_{\Sigma_2-\Sigma_1} H^1(I_q, A)^{D_q}$$

is also exact. Injectivity of the first map follows from the fact that $\mathbf{T}_{\Sigma_2} \longrightarrow \mathbf{T}_{\Sigma_1}$, and hence, $\Phi_{\mathbf{T}_{\Sigma_2}} \longrightarrow \Phi_{\mathbf{T}_{\Sigma_1}}$, is surjective.

Exactness at the second stage asserts that a modular deformation of level N_{Σ_2} which is unramified at the primes in $\Sigma_2 - \Sigma_1$, already arises from a modular form of level N_{Σ_1}. This is very much in the spirit of Ribet's celebrated "lowering the level" result [Ri2]. (Ribet only concerns himself with mod p modular forms, whereas what is needed here is some version for modular forms mod p^k.)

Say that a prime q is *good* if

$$(q+1)^2 - a_q^2 \neq 0 \pmod{5}.$$

Proposition 4.2 *(Wiles) If $\Sigma_2 - \Sigma_1$ consists only of good primes, then the sequence*

$$0 \longrightarrow S^{mod}_{\Sigma_1}(\mathbf{Q}, A) \longrightarrow S^{mod}_{\Sigma_2}(\mathbf{Q}, A) \longrightarrow \oplus_{\Sigma_2-\Sigma_1} H^1(I_q, A)^{D_q}$$

is exact.

This proposition is the main technical ingredient in Wiles' proof of the inequality (10). The proof relies in an essential way on the work of Ribet and on its subsequent refinements and extensions. For details, the reader may consult [Wi].

Now, let q be a good prime. In our particularly simple situation, we know that $S^{mod}_{\emptyset}(\mathbf{Q}, A) = 0$, and hence prop. 4.2 gives an injective map

$$S^{mod}_{\{q\}}(\mathbf{Q}, A) \hookrightarrow H^1(I_q, A)^{D_q}.$$

On the other hand

$$\#S^{mod}_{\{q\}}(\mathbf{Q}, A) = \#\Phi_{\mathbf{T}_{\{q\}}} \geq \#(\mathbf{Z}_5/\eta_{\mathbf{T}_{\{q\}}}) = [(q-1)((q+1)^2 - a_q^2)],$$

where the first inequality follows from equation (6), and the last equality is thm. 3.2. Hence by using equation (15) and counting orders, we find that the map $S^{mod}_{\{q\}}(\mathbf{Q}, A) \longrightarrow H^1(I_q, A)^{D_q}$ is an isomorphism. Hence, so is the map $S^{mod}_{\{q\}}(\mathbf{Q}, A)_5 \longrightarrow H^1(I_q, A)_5^{D_q}$. Since $S^{mod}_{\{q\}}(\mathbf{Q}, A)_5$ injects into $S_{\{q\}}(\mathbf{Q}, A)_5$, and since $S_{\{q\}}(\mathbf{Q}, A_5)$ surjects onto $S_{\{q\}}(\mathbf{Q}, A)_5$, we have shown:

Lemma 4.3 *If q is a good prime, then the map*

$$S_{\{q\}}(\mathbf{Q}, A_5) \longrightarrow H^1(I_q, A_5)^{D_q} = \hom_{D_q}(I_q, A_5)$$

is surjective.

4.2.2 Local Tate duality

In this section (and this section only!) let A be a *finite* $G_{\mathbf{Q}}$-module of p-power order, and let $A^* = \hom(A, G_m)$ be the Kummer dual of A, equipped with the natural Galois action.

Proposition 4.4 *(Tate) For all primes q cup-product induces a canonical non-degenerate pairing*

$$\langle\ ,\ \rangle_q : H^1(\mathbf{Q}_q, A) \times H^1(\mathbf{Q}_q, A^*) \longrightarrow \mathbf{Q}_p/\mathbf{Z}_p.$$

The local Tate pairing is defined by composing the natural maps

$$H^1(\mathbf{Q}_q, A) \times H^1(\mathbf{Q}_q, A^*) \xrightarrow{\cup} H^2(\mathbf{Q}_q, A \otimes A^*) = H^2(\mathbf{Q}_q, G_m) = \mathbf{Q}/\mathbf{Z},$$

where the first is given by cup product and the last equality follows from local class field theory [CF].

A simple description of the local Tate pairing can be given when $q \neq p$ and the module A is unramified at q. In this case, the inflation-restriction sequence gives an exact sequence:

$$0 \longrightarrow H^1(\mathbf{Q}_q^{nr}/\mathbf{Q}_q, A) \longrightarrow H^1(\mathbf{Q}_q, A) \longrightarrow H^1(I_q, A)^{D_q} \longrightarrow 0.$$

Since A is unramified, the term on the left can be identified with the module A_{D_q} of D_q-coinvariants of A, and the term on the right can be identified with

$$\hom(I_q, A)^{D_q} = A(-1)^{D_q}.$$

One can show that the submodules $A_{D_q} \subset H^1(\mathbf{Q}_q, A)$ and $A^*_{D_q} \subset H^1(\mathbf{Q}_q, A^*)$ are the orthogonal complements of each other under the local Tate pairing, and that the induced pairings on $A_{D_q} \times A^*(-1)^{D_q}$ and $A^*_{D_q} \times A(-1)^{D_q}$ are the obvious ones.

If a and $a^{'}$ are classes in $H^1(\mathbf{Q}, A)$ and $H^1(\mathbf{Q}, A^*)$ respectively, let a_v and $a^{'}_v$ denote their images in the local groups $H^1(\mathbf{Q}_v, A)$ and $H^1(\mathbf{Q}_v, A^*)$.

Proposition 4.5 *(Tate) For all a and $a^{'}$, we have*

$$\sum_v \langle a_v, a_v^{'} \rangle_v = 0.$$

Proof: We have

$$\langle a, a^{'} \rangle_v = \mathrm{inv}_v(a \cup a^{'}),$$

where $a \cup a^{'} \in H^2(\mathbf{Q}, G_m)$ is an element in the global Brauer group of $\mathbf{Q}$. Hence the proposition follows directly from the global reciprocity law of class field theory [CF] which states that

$$\sum_v \mathrm{inv}_v(x) = 0, \quad \forall x \in H^2(\mathbf{Q}, G_m).$$

Now, let $J_q \subset H^1(\mathbf{Q}_q, A)$ be a choice of subgroups of the local cohomology groups, satisfying

$$J_q = H^1(D_q/I_q, A^{I_q}) \quad \text{for almost all } q,$$

and for any such system of (J_q), define the generalized Selmer group:

$$S_{(J_q)}(\mathbf{Q}, A) =:= \{ s \in H^1(\mathbf{Q}, A) \quad | \quad s_q \in J_q \ \forall q \ \}.$$

If we define $J_q^* \subset H^1(\mathbf{Q}_q, A^*)$ to be the orthogonal submodules under the local Tate pairings, then J_q^* also satisfies the condition:

$$J_q^* = H^1(D_q/I_q, A^{*I_q}) \quad \text{for almost all } q.$$

We define the *dual Selmer group* of $S_{(J_q)}(\mathbf{Q}, A)$ to be the Selmer group $S_{(J_q^*)}(\mathbf{Q}, A^*)$.

The following Euler characteristic formula compares the orders of the Selmer group $S_{(J_q)}(\mathbf{Q}, A)$ and its dual $S_{(J_q^*)}(\mathbf{Q}, A^*)$.

Proposition 4.6 *The Selmer groups $S_{(J_q)}(\mathbf{Q}, A)$ and $S_{(J_q^*)}(\mathbf{Q}, A^*)$ are finite, and*

$$\frac{\#S_{(J_q)}(\mathbf{Q}, A)}{\#S_{(J_q^*)}(\mathbf{Q}, A^*)} = \frac{\#H^0(\mathbf{Q}, A)}{\#H^0(\mathbf{Q}, A^*)} h_\infty \prod_q h_q,$$

where

$$h_\infty = \#H^0(\mathbf{R}, A^*), \quad h_q = \#H^0(\mathbf{Q}_q, A^*)/[H^1(\mathbf{Q}_q, A) : J_q].$$

For a proof of the general formula, see [Gre]. Note that when A is unramified at $q \neq p$, then

$$J_q = H^1(D_q/I_q, A) \Rightarrow \#H^0(D_q/I_q, A^*) = [H^1(\mathbf{Q}_q, A) : J_q],$$

and $h_q = 1$. Hence, $h_q = 1$ for almost all q, so that the above product makes sense.

4.2.3 Bounding $S_\emptyset(\mathbf{Q}, A_5^*)$

Define $S_\Sigma(\mathbf{Q}, A_5^*)$ to be the dual Selmer group of $S_\Sigma(\mathbf{Q}, A_5)$, using the general construction of the previous section. Our goal in this section is to show

Proposition 4.7 $\#S_\emptyset(\mathbf{Q}, A_5^*) = 1$.

Proof: We begin by showing that any $s \in S_\emptyset(\mathbf{Q}, A_5^*)$ is locally trivial for many q. More precisely, if $s \in S_\emptyset(\mathbf{Q}, A_5^*)$, and q is any prime, let s_q be the image of s in $H^1(\mathbf{Q}_q, A_5^*)$ by restriction. If q is not 5 or 17, then s_q belongs to $H^1(D_q/I_q, A_5^*)$.

Lemma 4.8 *If s belongs to $S_\emptyset(\mathbf{Q}, A_5^*)$ and if q is a good prime, then $s_q = 0$.*

Proof: For all $\gamma \in S_{\{q\}}(\mathbf{Q}, A_5)$, we have, by prop. 4.5,

$$\sum_v \langle s_v, \gamma_v \rangle_v = 0. \tag{16}$$

But if $v \neq q$, then by definition $\langle s_v, \gamma_v \rangle_v = 0$. Hence, equation (16) reduces to:

$$\langle s_q, \gamma_q \rangle_q = 0, \quad \forall \gamma \in S_{\{q\}}(\mathbf{Q}, A_5).$$

By lemma 4.3 (which is the main actor in this proof), it follows that

$$\langle s_q, \alpha \rangle_q = 0, \quad \forall \alpha \in H^1(I_q, A_5)^{D_q}.$$

But this proves that $s_q = 0$, by the non-degeneracy of the local Tate pairing.

Let $K = \mathbf{Q}(A_5^*)(= \mathbf{Q}(A_5))$ be the splitting field of A_5^*, i.e., the smallest field through which $\bar{\rho}_0$ factors. By lemma 1.4, we have $G = \mathrm{Gal}(K/\mathbf{Q}) \simeq \mathbf{GL}_2(\mathbf{F}_5)$.

Lemma 4.9 *The restriction map*

$$H^1(\mathbf{Q}, A_5^*) \longrightarrow H^1(K, A_5^*)$$

is injective.

Proof: From the inflation-restriction sequence, the kernel of this map is

$$H^1(G, A_5^*) \simeq H^1(\mathbf{GL}_2(\mathbf{F}_5), \mathrm{Sym}^2),$$

where Sym^2 is the symmetric square of the standard two-dimensional representation of $\mathbf{GL}_2(\mathbf{F}_5)$. Let $Z \subset \mathbf{GL}_2(\mathbf{F}_5)$ denote the group of scalar matrices. The Hochschild-Serre spectral sequence

$$H^p(\mathbf{PGL}_2(\mathbf{F}_5), H^q(Z, \mathrm{Sym}^2)) \Rightarrow H^{p+q}(\mathbf{GL}_2(\mathbf{F}_5), \mathrm{Sym}^2)$$

shows that $H^p(\mathbf{GL}_2(\mathbf{F}_5), \mathrm{Sym}^2) = 0$ for all p, since the group Z is of order prime to 5.

We now come to the proof of prop. 4.7. Let s be an element of $S_\emptyset(\mathbf{Q}, A_5^*)$. We will show that $s = 0$. To show this, it suffices to show that the restriction $\bar{s}$ of s to $H^1(K, A_5^*)$ is zero, by lemma 4.9. Let L/K be the smallest extension of K such that $\bar{s} \in \hom(\mathrm{Gal}(\bar{K}/K), A_5^*)$ factors through $U = \mathrm{Gal}(L/K)$. If $\bar{s}$ is non-trivial, then the group U is isomorphic to $(\mathbf{Z}/5\mathbf{Z})^3$, and $G \simeq \mathbf{GL}_2(\mathbf{F}_5)$ acts on U by conjugation. Furthermore, U is isomorphic as a G-module to the symmetric square of the standard representation of $\mathbf{GL}_2(\mathbf{F}_5)$. Let $\Gamma = \mathrm{Gal}(L/\mathbf{Q})$, so that we have an exact sequence of finite groups:

$$0 \longrightarrow U \longrightarrow \Gamma \longrightarrow G \longrightarrow 0.$$

Now fix a $\tau \in \Gamma$ such that the image of τ in $G \simeq \mathbf{GL}_2(\mathbf{F}_5)$ is conjugate to a matrix of the form $\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$, and is of order 4 in Γ. (Such a τ always exists, since 4 is prime to 5.) Now let h be an arbitrary element of the group U, and choose a prime q of $\mathbf{Q}$ whose Frobenius element in $\mathrm{Gal}(L/\mathbf{Q}) = \Gamma$ is equal to $h\tau$. Such a prime exists, by the Chebotarev density theorem. Since $\bar{\rho}_0(\mathrm{Frob}_q)$ has eigenvalues 2 and 3, it follows that

$$a_q = 0 \pmod 5, \quad q = 1 \pmod 5,$$

so that $(q+1)^2 - a_q^2 = 4 \neq 0 \pmod 5$. Therefore q is a good prime. Hence we can apply lemma 4.8 to conclude that $s_q = 0$, so that, if $\mathcal{Q}$ is a prime of K above q,

$$\bar{s}(\mathrm{Frob}_{\mathcal{Q}}) = 0.$$

Since the extension $K_{\mathcal{Q}}/\mathbf{Q}_q$ has residue degree 4, we have

$$\mathrm{Frob}_{\mathcal{Q}} = (h\tau)^4 = h(\tau h \tau^{-1})(\tau^2 h \tau^{-2})(\tau^3 h \tau^{-3}) = h^+,$$

where h^+ is the projection of $h \in U$ to the $+1$-eigenspace, U^+, for the action of τ on U. Since $h \in U$ was arbitrary, it follows that $\bar{s}$ annihilates U^+. The eigenvalues of τ on U are -1, -1, and $+1$, and hence U^+ is non-trivial. Since U is an irreducilbe G-module, and $\bar{s}$ is G-equivariant, it follows that $\bar{s} = 0$, as was to be shown.

4.2.4 Bounding $S_\emptyset(\mathbf{Q}, A_5)$

Let $J_q \subset H^1(\mathbf{Q}_q, A_5)$ be the groups defined in sec. 3.1, but with A_5 replacing A, and let $J_q^* \subset H^1(\mathbf{Q}_q, A_5^*)$ denote the orhogonal complements under the local Tate pairings, so that

$$S_\Sigma(\mathbf{Q}, A_5) = S_{(J_q)}(\mathbf{Q}, A_5), \quad S_\Sigma(\mathbf{Q}, A_5^*) = S_{J_q^*}(\mathbf{Q}, A_5^*).$$

Applying proposition 4.6 in our situation, with $\Sigma = \emptyset$, we get:

$$\frac{\#S_\emptyset(\mathbf{Q}, A_5)}{\#S_\emptyset(\mathbf{Q}, A_5^*)} = h_\infty h_5 h_{17}, \tag{17}$$

where

$$h_\infty = \#H^0(\mathbf{R}, A_5^*), \quad h_q = \#H^0(\mathbf{Q}_q, A_5^*)/[H^1(\mathbf{Q}_q, A_5) : J_q].$$

Thus to compute $\#S_\emptyset(\mathbf{Q}, A_5)$ we are reduced to a series of local computations. We recall briefly the facts of local Galois cohomology that we will use. (For proofs of these, see [Sr2].)

1. If A is any finite module over $D_q = G_{\mathbf{Q}_q}$, let $h^i(A)$ denote $\#H^i(\mathbf{Q}_q, A)$. Then $h^i(A) = 1$ if $i > 2$, and the Euler characteristic of A is defined by

$$\chi(A) = \frac{h^0(A)h^2(A)}{h^1(A)}.$$

The Poitou-Tate formula for the local Euler characteristic is

$$\chi(A) = |\#A|_q,$$

where $|x|_q = q^{-ord_q(x)}$ is the usual q-adic valuation.
2. It is a direct consequence of the local Tate duality that $h^i(A) = h^{2-i}(A^*)$.

Lemma 4.10 $h_\infty = 25$.

Proof: Since $\bar{\rho}_0$ is odd, complex conjugation acts on A_5 with eigenvalues $1, -1, and -1$, and hence with eigenvalues $-1, 1, 1$ on A_5^*. The result follows.

Lemma 4.11 $h_{17} = 1$.

Proof: Let A_5^0 be the module $(A_{(17)}^0)_5$, so that J_17 is defined by the exact sequence

$$0 \longrightarrow J_{17} \longrightarrow H^1(\mathbf{Q}_{17}, A_5) \longrightarrow H^1(\mathbf{Q}_{17}, A_5/A_5^0).$$

The group $G_{\mathbf{Q}_{17}}$ acts on the line A_5^0 by the cyclotomic character χ. More generally, there is a filtration

$$A_5^0 \subset A_5^1 \subset A_5$$

with one dimensional quotients, and the action of $G_{\mathbf{Q}_{17}}$ on the sucessive quotients is by χ, 1, and χ^{-1}. Using this explicit description, one verifies that $h^0(A_5) = h^0(A_5^*) = 1$, that $h^0(A_5^0) = 1$, $h^0(A_5^{0*}) = 5$, etc.

From the long exact cohomology sequence associated to

$$0 \longrightarrow A_5^0 \longrightarrow A_5 \longrightarrow A_5/A_5^0 \longrightarrow 0,$$

we obtain

$$\begin{aligned}[H^1(\mathbf{Q}_{17}, A_5) : J_{17}] &= \frac{h^1(A_5/A_5^0)h^2(A_5)}{h^2(A_5^0)h^2(A_5/A_5^0)} = \frac{h^0(A_5/A_5^0)}{\chi(A_5/A_5^0)} \cdot \frac{h^0(A_5^*)}{h^0(A_5^{0*})} \\ &= \frac{5}{1} \times \frac{1}{5} = 1.\end{aligned}$$

Since $h^0(A_5^*) = 1$, one obtains $h_{17} = 1$, which proves the lemma.

Lemma 4.12 $h_5 \leq 1/25$.

Proof: Denote now by A_5^0 the module $(A_{(5)}^0)_5$, so that J_5 is defined by the exact sequence

$$0 \longrightarrow J_5 \longrightarrow H^1(\mathbf{Q}_5, A_5) \longrightarrow H^1(I_5, A_5/A_5^0).$$

The inertia group I_5 acts on the line A_5^0 by the cyclotomic character χ. More generally, there is a filtration

$$A_5^0 \subset A_5^1 \subset A_5$$

with one dimensional quotients, and the action of the decomposition group $G_{\mathbf{Q}_5}$ on the sucessive quotients is by $\chi\psi^{-2}$, 1, and $\chi^{-1}\psi^2$, where ψ is an unramified character of order 4. This follows from part 2 of lemma 1.4.

Using this explicit description, one verifies that $h^0(A_5) = h^0(A_5^*) = 1$, that $h^0(A_5^0) = h^0(A_5^0) = 1$, etc.

Now, consider the composite map

$$\phi : H^1(\mathbf{Q}_5, A_5) \xrightarrow{\phi_1} H^1(\mathbf{Q}_5, A_5/A_5^0) \xrightarrow{\phi_2} H^1(I_5, A_5/A_5^0).$$

Then we have

$$[H^1(\mathbf{Q}_5, A_5) : J_5] = \#\mathrm{Im}(\phi) \geq \#(\mathrm{Im}\phi_1)/\#(\ker\phi_2). \tag{18}$$

By the inflation restriction sequence,

$$\ker\phi_2 = H^1(D_5/I_5, (A_5/A_5^0)^{I_5}) = ((A_5/A_5^0)^{I_5})_{D_5/I_5},$$

so that

$$\#\ker\phi_2 = h^0(A_5/A_5^0) = 5. \tag{19}$$

On the other hand, by the same argument as in the proof of lemma 4.11, we find:

$$\#\mathrm{Im}\phi_1 = \frac{h^0(A_5/A_5^0)}{\chi(A_5/A_5^0)} \cdot \frac{h^0(A_5^*)}{h^0(A_5^{0*})} = \frac{5}{25^{-1}} \times \frac{1}{1} = 125.$$

(Now, of course, $h^i(M)$ denotes $\#H^i(\mathbf{Q}_5, M)$ instead of $\#H^i(\mathbf{Q}_{17}, M)$!) Hence, by eqn. (18) and (19),

$$[H^1(\mathbf{Q}_5, A_5) : J_5] \geq 25.$$

Since $h^0(A_5^*) = 1$, one obtains $h_5 \leq 1/25$, which proves the lemma.

Corollary 4.13 $\#S_\emptyset(\mathbf{Q}, A_5) = \#S_\emptyset(\mathbf{Q}, A_5^*) = 1$.

Proof: Combining lemmas 4.10, 4.11, and 4.12 yields the inequality

$$h_\infty h_5 h_{17} \leq 1.$$

The corollary now follows from eqn. (17). This finishes the proof of inequality (10), and hence of thm. 1.2.

5 Wiles' general strategy

We conclude by briefly mentionning Wiles' strategy for proving the Shimura-Taniyama conjecture for a large class of elliptic curves, containing all of the semi-stable ones. This can also be found in other surveys, eg. [RS], [Gou].
1. If E is an arbitrary (semistable) elliptic curve over $\mathbf{Q}$, to get the machinery of Wiles going one needs to know that E_p is modular for some prime p. It was conjectured by Serre [Sr3] that *any* odd representation with values in $\mathbf{GL}_2(\mathbf{F}_p)$ arises from a modular form with apropriate weight, character and level, but this conjecture seems quite difficult and is very much an open problem, even for $p = 5$. For $p = 3$, however, something very fortunate occurs: the image of the mod 3 representation is isomorphic to a double cover of S_4 which is a finite solvable subgroup of $\mathbf{GL}_2(\mathbf{C})$. A result of Langlands [La] and Tunnell [Tu] then guarantees that ρ_3 arises from a modular form (mod 3) of weight 1. By multiplying by an appropriate Eisenstein series, one gets a form of weight 2 mod 3 which gives the corresponding representation. Finally, the work of Ribet and others allows one to then show that there is a form of "minimal" level given by the recipe explained in the conjectures of [Sr3]. (See for example [Di].) Using this minimal level, Wiles constructs the Hecke ring $\mathbf{T}_\emptyset$ and the base point map $\mathbf{T}_\emptyset \longrightarrow \mathcal{O}$, where $\mathcal{O}$ is a local ring with residue field $\mathbf{F}_3$. He also gets rings $\mathbf{T}_\Sigma$ for sets of primes Σ by increasing the level.

2. Wiles then needs to construct a deformation ring R_Σ which "captures" the 3-adic representation associated to E. This leads him to consider a variety of deformation problems, corresponding to the cases where E is good ordinary, multiplicative, or supersingular at 3. In the ordinary and multiplicative cases, the deformation given by $Ta_3(E)$ is *ordinary*, and the existence of

the universal rings R_Σ is shown by Mazur and Tilouine [MT]. When E is supersingular at 3 one needs to consider flat deformations; here one makes essential use of the work of Ramakrishna [Ra], where the existence and key properties of the universal deformation ring are established.

3. Wiles then needs to show that the map $R_\Sigma \longrightarrow \mathbf{T}_\Sigma$ is an isomorphism for all sets Σ. Using the ideas explained in this report, he succeeds in reducing this isomorphism to the inequality

$$length(S_\emptyset(\mathbf{Q}, A)) \leq length(\mathcal{O}/\eta_{\mathbf{T}_\emptyset}\mathcal{O}), \tag{20}$$

for the ring $\mathbf{T}_\emptyset$ that arises in the minimal deformation problem. At the time of writing, this inequality has not yet been established in sufficient generality, although it can be established in many specific instances thanks to the ideas of Wiles and Flach, as we have explained.

4. The inequality (20) (for $p = 3$) is sufficient to show that any semi-stable elliptic curve such that E_3 is absolutely irreducible is modular. (N. Elkies has shown that this fact alone is enough to prove that all of the Frey curves, whose modularity implies Fermat's Last Theorem, satisfy the Shimura Taniyama conjecture.) By an elementary and ingenious argument, Wiles shows that *all* semistable curves are modular, assuming (20), as follows: If the mod 3 representation ρ_3 arising from E is reducible, Wiles constructs an auxiliary curve $E^{'}$ such that $E^{'}_3$ is absolutely irreducible as a Galois module, and E_5 is isomorphic to $E^{'}_5$ as Galois modules. The key point here is that the modular curve $X(5)$ is of genus 0, so that there is a plentiful supply of curves $E^{'}$ satisfying $E^{'}_5 \simeq E_5$. Now, applying his argument to the irreducible representation attached to $E^{'}_3$, Wiles shows that $E^{'}$ is modular. Hence, so is $E^{'}_5$, and therefore E_5. By an analysis of the rational points on $X_0(15)$, one knows that E_5 must be irreducible; hence, Wiles can apply his analysis, again assuming equation (20) with $p = 5$, to conclude that E is modular.

References

[BK] Bloch, S., and Kato, K., *L-functions and Tamagawa numbers of motives*, in: The Grothendieck Festschrift I, Progress in Math. 86, Birkhäuser, Boston, Basel, Berlin, 1990, p. 333-400.

[Ca1] H. Carayol, *Sur les représentations galoisiennes modulo l attachées aux formes modulaires*, Duke Math. J. **59** (1989), 785–801.

[Ca2] Carayol, H. *Formes modulaires et représentations galoisiennes à valeurs dans un anneau local complet*, in "p-adic monodromy and the Birch-Swinnerton-Dyer conjecture", (eds. B. Mazur and G. Stevens), Contemp. Math. 165 (1994).

[CF] Cassels, J.W.S., Frohlich, A., Algebraic number theory, Academic Press, London (1967).

[Co] Cox, D., *Introduction to Fermat's Last Theorem*, American Math. Monthly, 101, 3-14.

[Cr] Cremona, J.E., *Algorithms for modular elliptic curves*, Cambridge University Press, 1992.

[Di] F. Diamond, *The refined conjecture of Serre*, to appear.

[Ed] Edixhoven, B., *The weight in Serre's conjectures on modular forms*, Invent. Math. 109 (1992), 563-594.

[Fa1] Faltings, G., *p-adic Hodge theory*, J. of the A. M. S. **1** (1988) 255–299.

[Fa2] Faltings, G., *Crystalline cohomology and p-adic Galois representations*, in Algebraic analysis, geometry and number theory. Proceedings of the JAMI Inaugural Conference, J. I. Igusa, ed., Johns Hopkins University Press, Baltimore (1989) 25–80.

[Fl] Flach, M., *A finiteness theorem for the symmetric square of an elliptic curve*, Invent. math. **109** (1992) 307–327.

[Gou] Gouvêa, F.Q., *"A Marvelous Proof"*, Amer. Math. Monthly, pp. 203-221, March 1994.

[Gre] Greenberg, R., *Iwasawa theory for p-adic representations*, in Algebraic Number Theory in honor of K. Iwasawa, Adv. Studies in Pure Math. 17, pp. 97-138.

[Gro] Gross, B., *Kolyvagin's work on modular elliptic curves*, in L-functions and arithmetic, London Math. Soc. Lecture Notes **153** (1991) 235–256.

[Ka] K. Kato, Iwasawa theory and p-adic Hodge theory, manuscript.

[Kn] Knapp, A.W., *Elliptic curves*, Princeton Universiy Press, 1992.

[Ko1] Kolyvagin, V. A., *Finiteness of $E(\mathbf{Q})$ and sha$(E/\mathbf{Q})$ for a class of Weil curves*, Izv. Akad. Nauk, SSSR 52, (1988).

[Ko] Kolyvagin, V. A., *Euler systems*, in The Grothendieck Festschrift (Vol. II), P. Cartier, et. al., eds., Birkhäuser, Boston (1990) 435–483.

[La] Langlands, R., Base change for $GL(2)$, Ann. of Math. Studies **96**, Princeton University Press, Princeton (1980).

[Le] Lenstra, H.W., *Complete intersections and Gorenstein rings*, Number theory seminar, Berkeley, September 22 & 29, 1993.

[Mat] Matsumura, H., *Commutative ring theory*, Cambridge University Press, Cambridge,1986.

[Mz1] Mazur, B. *Modular curves and the Eisenstein ideal*, Publ. Math. IHES 47, 33-186 (1977)

[Mz2] Mazur, B., *Deforming Galois representations*, in Galois groups over **Q**, Y. Ihara, K. Ribet, J-P. Serre, eds., MSRI publications **16**, Springer-Verlag, New York (1989) 385–437.

[MRi] Mazur, B. and Ribet, K. *Two-dimensional representations in the arithmetic of modular curves.* To appear

[MRo] Mazur, B., Roberts, L., *Local Euler characteristics*, Invent. Math. 9 (1970), 201-234.

[MT] Mazur, B., Tilouine, J., *Représentations galoisiennes, différentielles de Kähler et "conjectures principales"*, Publ. Math. IHES **71** (1990) 65–103.

[Mi] J.S. Milne, Arithmetic duality theorems. Perspectives in mathematics. Academic Press, 1986.

[MF] Birch, B., Kuyk, W., eds., Modular functions of one variable IV, vol. 476, Springer-Verlag, New York (1975) 74–144.

[Mu] Murty, M.R., *Selberg's conjectures and Artin L-functions*, Bulletin of the A.M.S., to appear.

[Ne] Nekovář, J., *Kolyvagin's method for Chow groups of Kuga-Sato varieties*, Invent. math. 107, (1992) 99-125.

[Ra] Ramakrishna, R., *On a variation of Mazur's deformation functor*, Compositio Math. **87** (1993) 269-286.

[Ri1] Ribet, K., *Congruence relations between modular forms*, Proceedings of the International Congress of Mathematicians, Warsaw, August 16-24, pp. 503-514, 1983.

[Ri2] Ribet, K., *On modular representations of Gal*($\bar{\mathbf{Q}}/\mathbf{Q}$) *arising from modular forms*, Invent. Math. 100, 431-476 (1990).

[Ri3] Ribet, K. *Report on mod ℓ representations of Gal*($\bar{\mathbf{Q}}/\mathbf{Q}$), preprint.

[Ri4] Ribet, K., *From the Taniyama-Shimura conjecture to Fermat's Last Theorem*, Annales de la Faculté des Sciences de Toulouse.

[RH] Ribet, K., and Hayes, B., *Fermat's Last Theorem and Modern Arithmetic*, American Scientist, Vo. 82, March-April 1994, pp. 144-156.

[RS] Rubin, K., and Silverberg, A., *A report on Wiles' Cambridge lectures*, Bulletin of the AMS, to appear.

[Ru] Rubin, K., *The "main conjectures" of Iwasawa theory for imaginary quadratic fields*, Invent. Math. 103 (1991) 25-68.

[Sr1] Serre, J.-P., *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Inv. Math. 15, 259-331 (1972).

[Sr2] Serre, J-P., *Cohomologie Galoisienne*, Lecture Notes in Math., **5**, Springer Verlag, 1973.

[Sr3] Serre, J.-P., *Sur les représentations modulaires de degré* 2 *de Gal*($\bar{\mathbf{Q}}/\mathbf{Q}$), Duke Math. J. Vol. 54, no. 1, 179-230 (1987).

[ST] Serre, J.-P., and Tate, J., *Good reduction of abelian varieties*, Ann. of Math. 88, 492-517 (1968)

[Sh] Shimura, G., *Introduction to the arithmetic theory of automorphic forms*, Princeton University Press.

[St] Sturm, J., *Special values of zeta-functions, and Eisenstein series of half integral weight*, American Journal of Mathematics, Vol. 102, 1980.

[Th] Thaine, F., *On the ideal class groups of real abelian number fields*, Ann. of Math. **128** (1988) 1–18.

[Tu] Tunnell, J., *Artin's conjecture for representations of octahedral type*, Bull. A.M.S. **5** (1981) 173–175.

[Wi] Wiles, A., Manuscript, to appear in Inv. Math.