Modular elliptic curves
and
Fermat's Last Theorem

By ANDREW WILES*

For Nada, Clare, Kate and Olivia

\[ \text{Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos, et generaliter nullam in infinitum ultra quadratum potestatem in duos ejusdem nominis fas est dividere: cujus rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.} \]

Pierre de Fermat

Introduction

An elliptic curve over \( \mathbb{Q} \) is said to be modular if it has a finite covering by a modular curve of the form \( X_0(N) \). Any such elliptic curve has the property that its Hasse-Weil zeta function has an analytic continuation and satisfies a functional equation of the standard type. If an elliptic curve over \( \mathbb{Q} \) with a given \( j \)-invariant is modular then it is easy to see that all elliptic curves with the same \( j \)-invariant are modular (in which case we say that the \( j \)-invariant is modular). A well-known conjecture which grew out of the work of Shimura and Taniyama in the 1950’s and 1960’s asserts that every elliptic curve over \( \mathbb{Q} \) is modular. However, it only became widely known through its publication in a paper of Weil in 1967 [We] (as an exercise for the interested reader!), in which, moreover, Weil gave conceptual evidence for the conjecture. Although it had been numerically verified in many cases, prior to the results described in this paper it had only been known that finitely many \( j \) invariants were modular.

In 1985 Frey made the remarkable observation that this conjecture should imply Fermat’s Last Theorem. The precise mechanism relating the two was formulated by Serre as the \( \varepsilon \)-conjecture and this was then proved by Ribet in the summer of 1986. Ribet’s result only requires one to prove the conjecture for semistable elliptic curves in order to deduce Fermat’s Last Theorem.

*The work on this paper was supported by an NSF grant.
The same results hold if the image of the projective representation \( \hat{\rho}_0 \) associated to \( \rho_0 \) is isomorphic to \( A_4, S_4 \) or \( A_5 \).

Proof. (i) Let \( G = \text{im} \rho_0 \) and let \( Z \) denote the center of \( G \). Then we have a surjection \( G' \rightarrow (G/Z)' \) where the ' denotes the derived group. By Dickson's classification of the subgroups of \( \text{GL}_2(k) \) containing an element of order \( p \), \( (G/Z) \) is isomorphic to \( \text{PGL}_2(k') \) or \( \text{PSL}_2(k') \) for some finite field \( k' \) of characteristic \( p \) or possibly to \( A_5 \) when \( p = 3 \), cf. [Di, p. 260]. In each case we can find, and then lift to \( G' \), an element of order \( m \) with \( (m, p) = 1 \) and \( m \geq 3 \), except possibly in the case \( p = 3 \) and \( \text{PSL}_2(F_3) \approx A_4 \) or \( \text{PGL}_2(F_3) \approx S_4 \).

However in these cases \( (G/Z)' \) has order divisible by 4 so the 2-Sylow subgroup of \( G' \) has order greater than 2. Since it has at most one element of exact order 2 (the eigenvalues would both be \(-1\) since it is in the kernel of the determinant and hence the element would be \(-1\)) it must also have an element of order 4.

The argument in the \( A_4, S_4 \) and \( A_5 \) cases is similar.

(ii) Since \( \rho_0 \) is assumed absolutely irreducible, \( G = \text{im} \rho_0 \) has no fixed line. We claim that the same then holds for the derived group \( G' \). For otherwise since \( G' \subseteq G \) we could obtain a second fixed line by taking \( (gz) \) where \( (2^c) \) is the original fixed line and \( g \) is a suitable element of \( G \). Thus \( G' \) would be contained in the group of diagonal matrices for a suitable basis and either it would be central in which case \( G \) would be abelian or its normalizer in \( \text{GL}_2(k) \), and hence also \( G \), would have order prime to \( p \). Since neither of these possibilities is allowed, \( G' \) has no fixed line.

By Dickson's classification of the subgroups of \( \text{GL}_2(k) \) containing an element of order \( p \) the image of \( \text{im} \rho_0 \) in \( \text{PGL}_2(k') \) or \( \text{PSL}_2(k') \) for some finite field \( k' \) of characteristic \( p \) or possibly to \( A_5 \) when \( p = 3 \). The only one of these with a quotient group of order \( p \) is \( \text{PSL}_2(F_3) \) when \( p = 3 \). It follows that \( p \mid [G : G'] \) except in this one case which we treat separately. So assuming now that \( p \mid [G : G'] \) we see that \( G' \) contains a non-trivial unipotent element \( u \). Since \( G' \) has no fixed line there must be another noncommuting unipotent element \( v \) in \( G' \). Pick a basis for \( \rho_0 \langle G' \rangle \) consisting of their fixed vectors. Then let \( \tau \) be an element of \( \text{Gal}(\mathbb{Q}_{2}/\mathbb{Q}) \) for which the image of \( \rho_0(\tau) \) in \( G/G' \) is prescribed and let \( \rho_0(\tau) = \left( \begin{array}{cc} \alpha & \beta \\ \gamma & \delta \end{array} \right) \). Then

\[
\delta = \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) \left( \begin{array}{c} 1 \\ \tau \end{array} \right) \left( \begin{array}{c} 1 \\ \tau \beta \end{array} \right)
\]

has \( \det(\delta) = \det \rho_0(\tau) \) and trace \( \delta = a\alpha + c + b\beta + a + d \). Since \( p \geq 3 \) we can choose this trace to avoid any two given values (by varying \( a \)) unless \( r\alpha \beta + c = 0 \) for all \( r \). But \( r\alpha \beta + c \) cannot be zero for all \( r \) as otherwise \( \alpha = c = 0 \). So we can find a \( \tau \) for which the ratio of the eigenvalues is not \( \omega(\tau) \), \( \det(\delta) \) being, of course, fixed.
Now suppose that \(\text{im} \rho_0\) does not have order divisible by \(p\) but that the associated projective representation \(\overline{\rho}_0\) has image isomorphic to \(S_4\) or \(A_5\), so necessarily \(p \neq 3\). Pick an element \(\tau\) such that the image of \(\rho_0(\tau)\) in \(G/G'\) is any prescribed class. Since this fixes both \(\det \rho_0(\tau)\) and \(\omega(\tau)\) we have to show that we can avoid at most two particular values of the trace for \(\tau\). To achieve this we can adapt our first choice of \(\tau\) by multiplying by any element of \(G'\). So pick \(\sigma \in G'\) as in (i) which we can assume in these two cases has order 3. Pick a basis for \(\rho_0\), by extending scalars if necessary, so that \(D \leftrightarrow (\alpha, \alpha^{-1})\). Then one easily checks that if \(\rho_0(\tau) = (\sigma \, 1)\) we cannot have the traces of all of \(\tau, \sigma \tau\) and \(\sigma^2 \tau\) lying in a set of the form \(\{ \alpha \tau \} \) unless \(a = d = 0\). However we can ensure that \(\rho_0(\tau)\) does not satisfy this by first multiplying \(\tau\) by a suitable element of \(G'\) since \(G'\) is not contained in the diagonal matrices (it is not abelian).

In the \(A_4\) case, and in the \(\text{PSL}_2(\mathbb{F}_3) \cong A_4\) case when \(p = 3\), we use a different argument. In both cases we find that the 2-Sylow subgroup of \(G/G'\) is generated by an element \(z\) in the centre of \(G\). Either a power of \(z\) is a suitable candidate for \(\rho_0(\sigma)\) or else we must multiply the power of \(z\) by an element of \(G'\), the ratio of whose eigenvalues is not equal to 1. Such an element exists because in \(G'\) the only possible elements without this property are \(\{ \pi \}\) (such elements necessarily have determinant 1 and order prime to \(p\)) and we know that \#\((G') > 2\) as was noted in the proof of part (i).

Remark. By a well-known result on the finite subgroups of \(\text{PGL}_2(\mathbb{F}_p)\) this lemma covers all \(\rho_0\) whose images are absolutely irreducible and for which \(\rho_0\) is not dihedral.

Let \(K_1\) be the splitting field of \(\rho_0\). Then we can view \(W_1\) and \(W_1^*\) as \(\text{Gal}(K_1(\mathbb{F}_p)/\mathbb{Q})\)-modules. We need to analyze their cohomology. Recall that we are assuming that \(\rho_0\) is absolutely irreducible. Let \(\overline{\rho}_0\) be the associated projective representation to \(\text{PGL}_2(k)\).

The following proposition is based on the computations in [CPS].

**Proposition 1.11.** Suppose that \(\rho_0\) is absolutely irreducible. Then
\[
H^1(K_1(\mathbb{F}_p)/\mathbb{Q}, W_1^*) = 0.
\]

**Proof.** If the image of \(\rho_0\) has order prime to \(p\) the lemma is trivial. The subgroups of \(\text{GL}_2(k)\) containing an element of order \(p\) which are not contained in a Borel subgroup have been classified by Dickson [Di, §260] or [Hu, II.8.27]. Their images inside \(\text{PGL}_2(k')\) where \(k'\) is the quadratic extension of \(k\) are conjugate to \(\text{PGL}_2(F)\) or \(\text{PSL}_2(F)\) for some subfield \(F\) of \(k'\), or they are isomorphic to one of the exceptional groups \(A_4, S_4, A_5\).

Assume then that the cohomology group \(H^1(K_1(\mathbb{F}_p)/\mathbb{Q}, W_1^*) \neq 0\). Then by considering the inflation-restriction sequence with respect to the normal
subgroup \( \text{Gal}(K_1(\zeta_p)/K_1) \) we see that \( \zeta_p \in K_1 \). Next, since the representation is (absolutely) irreducible, the center \( Z \) of \( \text{Gal}(K_1/Q) \) is contained in the diagonal matrices and so acts trivially on \( W_\lambda \). So by considering the inflation-restriction sequence with respect to \( Z \) we see that \( Z \) acts trivially on \( \zeta_p \) (and on \( W_\lambda^* \)). So \( \text{Gal}(Q(\zeta_p)/Q) \) is a quotient of \( \text{Gal}(K_1/Q)/Z \). This rules out all cases when \( p \neq 3 \), and when \( p = 3 \) we only have to consider the case where the image of the projective representation is isomorphic as a group to \( \text{PGL}_2(F) \) for some finite field of characteristic 3. (Note that \( S_3 \cong \text{PGL}_2(F) \).)

Extending scalars commutes with formation of duals and \( H^1 \), so we may assume without loss of generality \( F \subseteq k \). If \( p - 3 \) and \( \neq F \neq 3 \) then \( H^1(\text{PSL}_2(F), W_\lambda) = 0 \) by results of [CPS]. Then if \( \rho_0 \) is the projective representation associated to \( \rho_0 \) suppose that \( g^{-1} \text{im} \rho_0 g = \text{PGL}_2(F) \) and let \( H = g \text{PSL}_2(F)g^{-1} \). Then \( W_\lambda \cong W_\lambda^* \) over \( H \) and

\[
H^1(H, W_\lambda) \otimes \bar{\bar{F}} \cong H^1(g^{-1}Hg, g^{-1}(W_\lambda \otimes \bar{\bar{F}})) = 0.
\]

We deduce also that \( H^1(\text{im}\rho_0, W_\lambda^*) = 0 \).

Finally we consider the case where \( F = F_3 \). I am grateful to Taylor for the following argument. First we consider the action of \( \text{PSL}_2(F_3) \) on \( W_\lambda \) explicitly by considering the conjugation action on matrices \( \{A \in M_2(F_3) : \text{trace } A = 0\} \). One sees that no such matrix is fixed by all the elements of order 2. Hence \( H^1(\text{PSL}_2(F_3), W_\lambda) \cong H^1(Z/3, (W_\lambda)_{C_2 \times C_2}) = 0 \)

where \( C_2 \times C_2 \) denotes the normal subgroup of order 4 in \( \text{PSL}_2(F_3) \cong A_4 \). Next we verify that there is a unique copy of \( A_4 \) in \( \text{PGL}_2(F_3) \) up to conjugation. For suppose that \( A, B \in \text{GL}_2(F_3) \) are such that \( A^2 = B^2 = I \) with the images of \( A, B \) representing distinct nontrivial commuting elements of \( \text{PGL}_2(F_3) \). We can choose \( A = \left( \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \) by a suitable choice of basis, i.e., by a suitable conjugation. Then \( B \) is diagonal or antidiagonal as it commutes with \( A \) up to a scalar, and as \( B, A \) are distinct in \( \text{PGL}_2(F_3) \) we have \( B = \left( \begin{array}{cc} a & 0 \\ 0 & a^{-1} \end{array} \right) \) for some \( a \). By conjugating by a diagonal matrix (which does not change \( A \)) we can assume that \( a = 1 \). The group generated by \( \{A, B\} \) in \( \text{PGL}_2(F_3) \) is its own centralizer so it has index at most 6 in its normalizer \( N \). Since \( N/(A, B) \cong S_3 \) there is a unique subgroup of \( N \) in which \( (A, B) \) has index 3 whence the image of the embedding of \( A_4 \) in \( \text{PGL}_2(F_3) \) is indeed unique (up to conjugation). So arguing as in (1.18) by extending scalars we see that \( H^1(\text{im}\rho_0, W_\lambda^*) = 0 \) when \( F = F_3 \) also.

The following lemma was pointed out to me by Taylor. It permits most dihedral cases to be covered by the methods of Chapter 3 and [TW].

**Lemma 1.12.** Suppose that \( \rho_0 \) is absolutely irreducible and that

\( \rho_0 \) is dihedral (the case where the image is \( \mathbb{Z}/2 \times \mathbb{Z}/2 \) is allowed).
(b) \( \rho_0 \big|_L \) is absolutely irreducible where \( L = \mathbb{Q} \left( \sqrt{(-1)^{(p-1)/2}p} \right) \).

Then for any positive integer \( n \) and any irreducible Galois stable subspace \( X \) of \( W_\chi \otimes \bar{k} \) there exists an element \( \sigma \in \text{Gal}(\mathbb{Q}/\mathbb{Q}) \) such that

(i) \( \rho_0(\sigma) \neq 1 \),

(ii) \( \sigma \) fixes \( \mathbb{Q}(\zeta_p) \),

(iii) \( \sigma \) has an eigenvalue 1 on \( X \).

**Proof.** If \( \rho_0 \) is dihedral then \( \rho_0 \otimes \bar{k} = \text{Ind}_H^G \chi \) for some \( H \) of index 2 in \( G \), where \( G = \text{Gal}(\mathbb{K}_1/\mathbb{Q}) \). (As before, \( \mathbb{K}_1 \) is the splitting field of \( \rho_0 \).) Here \( H \) can be taken as the full inverse image of any of the normal subgroups of index 2 defining the dihedral group. Then \( W_\chi \otimes \bar{k} = \delta \oplus \text{Ind}_H^G(\chi/\chi') \), where \( \delta \) is the quadratic character \( G \to \mathbb{C}^\times \) and \( \chi' \) is the conjugate of \( \chi \) by any element of \( G \). Note that \( \chi \neq \chi' \) since \( H \) has nontrivial image in \( \text{PGL}_2(\bar{k}) \).

To find a \( \sigma \) such that \( b(\sigma) = 1 \) and conditions (i) and (ii) hold, observe that \( M(\zeta_p) \) is abelian where \( M \) is the quadratic field associated to \( \delta \). So conditions (i) and (ii) can be satisfied if \( \rho_0 \) is non-abelian. If \( \rho_0 \) is abelian (i.e., the image has the form \( \mathbb{Z}/2 \times \mathbb{Z}/2 \)), then we use hypothesis (b). If \( \text{Ind}_H^G(\chi/\chi') \) is reducible over \( \bar{k} \) then \( W_\chi \otimes \bar{k} \) is a sum of three distinct quadratic characters, none of which is the quadratic character associated to \( L \), and we can repeat the argument by changing the choice of \( H \) for the other two characters. If \( X = \text{Ind}_H^G(\chi/\chi') \otimes \bar{k} \) is absolutely irreducible then pick any \( \sigma \in G - H \). This satisfies (i) and can be made to satisfy (ii) if (b) holds. Finally, since \( \sigma \in G - H \) we see that \( \sigma \) has trace zero and \( \sigma^2 = 1 \) in its action on \( X \). Thus it has an eigenvalue equal to 1. \( \square \)

**Chapter 2**

In this chapter we study the Hecke rings. In the first section we recall some of the well-known properties of these rings and especially the Gorenstein property whose proof is rather technical, depending on a characteristic \( p \) version of the \( q \)-expansion principle. In the second section we compute the relations between the Hecke rings as the level is augmented. The purpose is to find the change in the \( \eta \)-invariant as the level increases.

In the third section we state the conjecture relating the deformation rings of Chapter 1 and the Hecke rings. Finally we end with the critical step of showing that if the conjecture is true at a minimal level then it is true at all levels. By the results of the appendix the conjecture is equivalent to the
that $S \cong T$, as otherwise $S \cong \ker \alpha \oplus \im \alpha$ is a nontrivial decomposition as $S$-modules, which contradicts $S$ being local.

\begin{proof}
\end{proof}

Remark. Lenstra has made an important improvement to this proposition by showing that replacing $\pi T$ by $J(\ann p)$ gives a criterion valid for all local $O$-algebras which are finite and free over $O$, thus without the Gorenstein hypothesis.

\section*{References}

\begin{thebibliography}{10}


\bibitem[Ca3]{Ca3} Formes modulaires et représentations galoisiennes $\ell$-adiques dans un anneau local complet. in \textit{$p$-adic Monodromy and the Birch-Swinnerton-Dyer Conjecture} (eds. B. Mazur and G. Stevens), Contemp. Math., vol. 165, 1994.


\bibitem[Co]{Co} R. Coleman, Division values in local fields, Invent. Math. 53 (1979), 91–116.


\bibitem[Dia]{Dia} F. Diamond, The refined conjecture of Serre, to appear in Proc. 1993 Hong Kong Conf. on Modular Forms and Elliptic Curves.


\end{thebibliography}
MODULAR ELLIPTIC CURVES AND FERMAT'S LAST THEOREM

[Dr] V. Deuring, Two-dimensional l-adic representations of the fundamental group of a curve over a finite field and automorphic forms on GL(2), Am. J. Math. 105 (1983), 85-114.


550


(Received October 14, 1994)