Compositio Mathematica **112:** 363–376, 1998.

A local analysis of congruences in the (p, p) case: Part I

CHANDRASHEKHAR KHARE
School of Mathematics, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India; e-mail: Shekhar@math.tifr.res.in

Received 2 August 1995; accepted in final form 12 May 1997

Abstract. Given an irreducible, modular, mod p representation ρ, we analyse the local components at p of newforms f which give rise to it.

Mathematics Subject Classifications (1991): 11F, 11R.

Key words: Congruences, mod p representations, automorphic representations, newforms.

Introduction

Suppose that

$$\rho: G_{\mathbb{Q}} \to GL_2(\bar{\mathbb{F}}_p)$$

is a continuous, odd, irreducible representation. Then Serre, cf. [S], conjectures that this representation is modular, i.e., arises from a newform $f \in S_k(\Gamma_1(M))$ for positive integers k and M. We *assume* that ρ is *modular*. In this is implicit the choice of embeddings $\iota_p: \bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_p$ and $\iota_\infty: \bar{\mathbb{Q}} \hookrightarrow \mathbb{C}$ which we fix. The choice of ι_p also fixes a place of $\bar{\mathbb{Q}}$ which we denote by $\mathfrak{p}$. We assume that $p \geqslant 3$.

It is interesting to study the behaviour of newforms which give rise to ρ. Thus we want to study the association:

$$\{f\} \to \rho,$$

where on the left hand side we consider newforms f which give rise to ρ.

Corresponding to f there is an automorphic representation of $GL_2(\mathbb{A}_{\mathbb{Q}})$ which we denote by $\pi(f)$. The above association was studied by Hida, Ribet, Carayol, Wiles and others, cf. for instance [H], [C] and [R]. In the work of Diamond–Taylor, cf. [DT], it is studied in terms of analysing the local components $\pi_\ell(f)$, at primes $\ell \neq p$, of the newforms f which figure in the above correspondence. Further the analysis is local. By this we mean the following. A determination of the possible admissible representations π_ℓ of $GL_2(\mathbb{Q}_\ell)$, such that there exists a newform f with the restriction of $\pi_\ell(f)$ to $GL_2(\mathbb{Z}_\ell)$ isomorphic to that of π_ℓ, and f gives rise to ρ, is carried out in terms of the restriction of ρ to a decomposition group at ℓ which we denote by ρ_ℓ (see Theorem 1 of [DT] for a more precise statement).

The aim of this paper, and its sequel, is to understand the local components at p, i.e., the nature of $\pi_p(f)$'s of newforms f which give rise to ρ. We refer to this as the (p, p) case.

In this paper we mainly restrict our attention to forms f of weight 2, such that $\pi_p(f)$ is either principal series (Section 1) or special (Section 2). This determination can be carried out by extensions of some of the methods in our earlier paper [K]. In Section 3 we present an example which illustrates the different nature of the theory of congruences in the (p, p) case. In Section 4 we sharpen the main theorem of [K] in one case.

The results proved here refine, in some instances, those in [K], where we had determined the coarser behaviour of the f's which give rise to ρ. Namely, there the p-part of the levels and nebentypes of newforms f which give rise to ρ, was analysed.

In a sequel to this paper, we will determine the nature of the $\pi_p(f)$'s of newforms f which give rise to ρ solely in terms of (the isomorphism type of) ρ_p, which denotes the restriction of ρ to D_p, the decomposition group at p fixed by the choice of the place $\mathfrak{p}$ above p, even allowing the case when $\pi_p(f)$ is supercuspidal, which we exclude here (but see Section 3). This will be done by a quite different method, which perhaps is conceptually more transparent. The main ingredients of the method are the following:

1. The proof of the 'weight part' of Serre's conjectures by Edixhoven, cf. [E].
2. Ribet's criterion for ρ to arise from a weight 2 newform which is special at p, cf. [R].
3. A version of a lemma of Carayol, cf. [C], in the setting of modular curves with the groups $GL_2(\mathbb{Z}/p^n)$ acting on them by automorphisms (similar to Lemma 9 in [DT]).
4. The classical fact that the isomorphism types of irreducible mod p representations of $GL_2(\mathbb{Z}/p^r)$ are given by $\mathbb{L}^n(\mathbb{F}_p) \otimes \chi^j$. The notation is as follows:

$\mathbb{L}^n(\mathbb{F}_p)$ is the symmetric n-th power of the standard two-dimensional representation over $\mathbb{F}_p$ of $GL_2(\mathbb{Z}/p^r)$. The character χ is the one-dimensional representation into $\mathbb{F}_p^*$ of $GL_2(\mathbb{Z}/p^r)$ given by the mod p reduction of the determinant character (which naturally takes values in $(\mathbb{Z}/p^r)^*$). The n's and j's are integers such that $0 \leqslant n \leqslant p-1$ and $0 \leqslant j \leqslant p-2$.

Using these facts the association

$$\pi_p(f) \to \rho_p$$

can be studied fully, where f runs through newforms which give rise to ρ.

But as said earlier, in this paper we content ourselves with proving what can be deduced from the methods of [K] (see Remark 13 of loc. cit.).

1. Principal Series liftings

Fix an irreducible modular mod p Galois representation ρ. In this section we analyse the newforms which give rise to ρ, and which are principal series at p.

In the theorem below, χ is the mod p cyclotomic character of $G_{\mathbb{Q}}$ giving the action on the p-th roots of unity. Given any mod p character of I_p, we may naturally associate to it a character of $\mathbb{Z}_p^*$ by local class field theory. We normalise the isomorphism so that an element $x \in \mathbb{Z}_p^*$ goes to the element σ_x of I_p, which is characterised by the fact that its action on μ, a root of unity of order a power of p, is given by $\mu \to \mu^{x^{-1}}$. Thus in the theorem below we associate to the character $\det(\rho)\chi^{1-k}$ (for some integer k) restricted to I_p, the corresponding character of $\mathbb{Z}_p^*$, which we denote by the same symbol.

THEOREM 1.1. *Let ρ be an irreducible modular mod p representation which arises from $S_k(\Gamma_1(Np^r))$ for an integer $N \geqslant 1$ with $(N,p) = 1$, and for some integers $k \geqslant 2$ and $r \geqslant 0$.*

1. *Let $(\varepsilon_1, \varepsilon_2)$ be a pair of characters of $\mathbb{Z}_p^*$ of finite conductor. Suppose there is a newform f of weight k which gives rise to ρ and $\pi_p(f)$ is a principal series of the form $\pi(\xi_1, \xi_2)$, with ξ_i quasicharacters of $\mathbb{Q}_p^*$ such that $\xi_i|\mathbb{Z}_p^* = \varepsilon_i$ for $i = 1, 2$. Then $\varepsilon_1\varepsilon_2$ as a mod $\mathfrak{p}$ character of $\mathbb{Z}_p^*$ is the fixed character* $\det(\rho)\chi^{1-k}$.

2. *Let $(\varepsilon_1, \varepsilon_2)$ be a pair of such characters of $\mathbb{Z}_p^*$, i.e., of finite conductors and such that the mod $\mathfrak{p}$ reduction of $\varepsilon_1\varepsilon_2$ is the character* $\det(\rho)\chi^{1-k}$. *Further assume that the order of $\varepsilon_1\varepsilon_2^{-1}$ is divisible by p. Then there exists a newform $f \in S_k(\Gamma_1(Mp^s))$, for $M|N$ and some non-negative integer s, with $\pi_p(f)$ a principal series of the form $\pi(\xi_1, \xi_2)$ for some quasicharacters ξ_i of $\mathbb{Q}_p^*$, with $\xi_i|\mathbb{Z}_p^* = \varepsilon_i$ for $i = 1, 2$, and such that f gives rise to ρ.*

Proof. (1) This follows by considering the nebentype ψ of $f \in S_k(\Gamma_0(Np^s), \psi)$ with $\pi_p(f)$ of the form in part 1, and using the relation that $\det(\rho)(\mathrm{Frob}_q) = \tilde{\psi}(q)\chi(q)^{k-1}$ if f gives rise to ρ (the tilde sign stands for reduction mod $\mathfrak{p}$) for almost all primes q.

(2) Assume that we are given a pair of characters $(\varepsilon_1, \varepsilon_2)$ as in 2 of the theorem. Let us write this pair as $(\omega^a\varepsilon_1', \omega^b\varepsilon_2')$ where ω is the Teichmuller character and ε_i' are characters of $\mathbb{Z}_p^*$ of orders powers of p.

We use Carayol's lemma, cf. [C], which applies because in the troublesome case of $p = 3$ (cf. [D]), $S_k(\Gamma_1(N)) = 0$, if $k \leqslant 4$ and $N \leqslant 4$. Thus from the weight part of Serre's conjecture, in the case of $p = 3$, the irreducibility of ρ implies $N > 4$, and thus $\Gamma_1(Np^r)$ will be torsion-free. (This remark is due to the referee.) Carayol's lemma, together with the well-known fact that ρ, as also any of its twists, arises from $S_k(\Gamma_1(Np^2))$ (see [R1] for this), implies that ρ arises from a newform g in $S_k(\Gamma_1(Np^2))$, such that $\pi_p(g)$ is principal series of the form $\pi(\zeta_1, \zeta_2)$. Here ζ_1 is a quasicharacter of conductor p^2 with $\zeta_1|\mathbb{Z}_p^* = \omega^c\varepsilon'', \varepsilon''$ is a character of conductor p^2 and order p, c is some integer, and ζ_2 is an unramified quasicharacter. Applying Carayol's lemma to $\rho \otimes \omega^{-b}$ and using the fact that $c \equiv a + b$ (mod

https://doi.org/10.1023/A:1000424731365 Published online by Cambridge University Press

$p-1$) (which follows from part 1), we see that $\rho \otimes \omega^{-b}$ arises from a weight k newform h such that $\pi_p(h)$ is principal series of the form $\pi(\xi_1', \xi_2')$, where $\xi_i' = 1, 2$ are quasicharacters with $\xi_1'|\mathbb{Z}_p^* = \omega^{a-b}\varepsilon_1'\varepsilon_2'^{-1}$ and ξ_2' unramified. We have crucially used the assumption that $\varepsilon_1\varepsilon_2^{-1}$ is a non-trivial character of $\mathbb{Z}_p^*$ of order divisible by p, as then we can apply the lemma of Carayol, where we work with the modular curve $X_1(Np^r)$ for $r \geqslant 2$. Note that the newform g above has level divisible by p^2. We refer the reader to the discussion in Remarks 8 and 11 of [K] for an explanation of how Carayol's lemma, applied in the setting of modular curves, yields 'principal series liftings' of ρ which we are using here. Succinctly, the point simply is that a newform f in $S_2(\Gamma_1(Np^r))$, such that its nebentype has conductor divisible by p^r, is principal series at p.

Now if we twist h by $\omega^b\varepsilon_2'$, we get a form f such that $\pi_p(f)$ is of the desired type, and f gives rise to ρ by construction. This proves part 2 of the theorem.

Note. The method used in part 2 of Theorem 1.1, namely to perform the operations of twisting, applying the lemma of Carayol and (un)twisting in succession, was already used in the proof of Theorem 5 of [K].

The analysis of principal series liftings by forms f in $S_k(\Gamma_1(Np^2))$ and where $\pi_p(f)$ is principal series of type $\pi(\xi_1, \xi_2)$ (where ξ_i are quasicharacters of $\mathbb{Q}_p^*$ of conductors dividing p), which is excluded in the above theorem, is more involved. We give below such analysis. We first consider the weight 2 case, and then deal with higher weights, using this analysis of the weight 2 case.

We first state necessary and sufficient conditions for ρ to arise from a newform in $S_2(\Gamma_1(Np))$. These were noted in Remark 9 of [K] to follow from Proposition 8.13 and 8.18 of [G], and the proof of the weight part of Serre's conjecture in [E]. The condition is that $\rho|I_p$ (I_p being the inertia group for the place $\mathfrak{p}$) be of one of the following forms:

1. $$\begin{pmatrix} \chi^a & * \\ 0 & 1 \end{pmatrix},$$

or

2. $$\begin{pmatrix} \psi^a & 0 \\ 0 & \psi'^a \end{pmatrix},$$

or

3. $$\begin{pmatrix} \chi & * \\ 0 & \chi^a \end{pmatrix},$$

where $1 \leqslant a \leqslant p-1$, χ is the mod $\mathfrak{p}$ cyclotomic character, and ψ and ψ' are the fundamental characters of level 2 of the tame quotient of I_p.

The principal series liftings of ρ of level dividing p^2 depend on the behaviour of $\rho|I_p$. In what follows, as a notational convenience, we will use the notation ξ_1

and ξ_2 for generic unramified quasi-characters of $\mathbb{Q}_p^*$. So for instance if ξ_1 occurs in two different places it does not mean that the same unramified quasi-character is intended!

We divide the analysis into two cases.

CASE 1. $\rho|I_p$ is semisimple

In this case $\rho|I_p$ is either of the form:

$$\begin{pmatrix} \chi^a & 0 \\ 0 & \chi^b \end{pmatrix}, \tag{A}$$

or

$$\begin{pmatrix} \psi^a\psi'^b & 0 \\ 0 & \psi^b\psi'^a \end{pmatrix}, \tag{B}$$

where in A, a and b are normalised so that $0 \leqslant a, b \leqslant p-2$, and in B the normalisation is that $0 \leqslant a < b \leqslant p-1$.

Then in case $\rho|I_p$ is of the form A, from what we have noted above, the only twists of ρ which will arise from $S_2(\Gamma_1(Np))$ are $\rho\otimes\chi^{-a}, \rho\otimes\chi^{-b}, \rho\otimes\chi^{1-a}$ and $\rho\otimes\chi^{1-b}$. In each of these cases the twisted form arises from a newform in $S_2(\Gamma_1(Np))$ such that it is principal series at p and is of the form $\pi(\omega^{b-a-1}\xi_1, \xi_2), \pi(\omega^{a-b-1}\xi_1, \xi_2), \pi(\omega^{b-a+1}\xi_1, \xi_2)$ and $\pi(\omega^{a-b+1}\xi_1, \xi_2)$, as follows from Remark 9 of [K], whose content we have recalled above. Thus we see that ρ arises from newforms in $S_2(\Gamma_1(Np)\cap\Gamma_0(p^2))$ which are principal series at p of the types $\pi(\omega^{b-1}\xi_1, \omega^a\xi_2)$ or $\pi(\omega^{a-1}\xi_1, \omega^b\xi_2)$, and these are the only possible types.

In the case when $\rho|I_p$ is of the form B, then again we may check that if $b \neq a+1$, the only twists of ρ which will arise from $S_2(\Gamma_1(Np))$ are $\rho\otimes\chi^{-a}$ and $\rho\otimes\chi^{1-b}$. In the case $b = a+1$, the only such twists is $\rho\otimes\chi^{-a}$ (see proof of Theorem 4.3 of [E]). From this we may see (as above; we skip the details) that ρ arises from newforms in $S_2(\Gamma_1(Np)\cap\Gamma_0(p^2))$ which are principal series at p of type $\pi(\omega^a\xi_1, \omega^{b-1}\xi_2)$, and this is the only possible type.

CASE 2. $\rho|I_p$ is non-semisimple

In this case $\rho|_p$ is of the form:

$$\begin{pmatrix} \chi^a & * \\ 0 & \chi^b \end{pmatrix}. \tag{C}$$

To harmonise our notations with those in [S], we assume that $1 \leqslant a \leqslant p-1$ and $0 \leqslant b \leqslant p-2$. We assume first that $a \neq b+1$. We see then that the only twists of ρ which will arise from a newform in $S_2(\Gamma_1(Np))$ which is principal series at p, are

$\rho \otimes \chi^{-b}$ and $\rho \otimes \chi^{1-a}$. These will arise from newforms which at p are principal series of the form $\pi(\omega^{a-b-1}\xi_1, \xi_2)$ and $\pi(\omega^{1+b-a}\xi_1, \xi_2)$. Thus we see that ρ will arise from newforms in $S_2(\Gamma_1(Np) \cap \Gamma_0(p^2))$ which are principal series at p of the type $\pi(\omega^{a-1}\xi_1, \omega^b\xi_2)$, and these are the only possible types.

In the case when $a = b+1$ things are slightly different. Namely, when $\rho \otimes \omega^{-b}$ is finite at p, then by a similar reasoning to the above (after using the fact that $\rho \otimes \omega^{-b}$ arises from $S_2(\Gamma_1(N))$ due to Mazur; see [E]), we see that ρ arises from a newform in $S_2(\Gamma_1(Np) \cap \Gamma_0(p^2))$ which is principal series at p of the type $\omega^b \otimes \pi(\xi_1, \xi_2)$ (note that as $a = b + 1$ this is of the same form as above), and this is the only possible type. When $\rho \otimes \omega^{-b}$ is not finite at p, we see by similar methods that ρ does not arise from any newform f in $S_2(\Gamma_1(Np) \cap \Gamma_0(p^2))$ such that $\pi_p(f)$ is principal series. Alternatively, we can deduce this more directly, as in the proof of case 3 of Theorem 3 of [K], by noting that such a newform f has the property that the corresponding mod p representation becomes finite at p after base changing to the ring of integers of a tamely ramified extension of $\mathbb{Q}_p$, which is not the case for ρ with $k(\rho) = p + 1$.

Summing up all this we have the following theorem:

THEOREM 1.2. *There exists a newform in $f \in S_2(\Gamma_1(Np^r))$ which gives rise to ρ, such that $\pi_p(f)$ is a principal series given by $\varepsilon \otimes \pi(\xi_1, \xi_2)$, where ξ_i are quasicharacters of $\mathbb{Q}_p^*$ such that $\xi_1|\mathbb{Z}_p^* = \varepsilon\omega^c$ and $\xi_2|\mathbb{Z}_p^* = \varepsilon\omega^d, \varepsilon$ a given Dirichlet character whose conductor and order are both powers of p, and some given integers c and d, iff the (unordered) pair $(c, d) \bmod p - 1$ is given by (with notation as above):*

1. *$(a, b - 1)$ or $(a - 1, b)$ if $\rho|I_p$ is semisimple, and of the form A above.*
2. *$(a, b - 1)$ if $\rho|I_p$ is semisimple, and of the form B above.*
3. *$(a - 1, b)$ if $\rho|I_p$ is not semisimple (is of the form C above), and no twist of ρ has Serre invariant $p + 1$.*
4. *No such c and d exists if some twist of ρ has Serre invariant $p + 1$.*

1.1. Principal Series Lifts of Level Np^2 for Higher Weights

Suppose ρ arised from $S_{k'}(\Gamma_1(N))$ for some weight $k' \geqslant 2$, and an integer N prime to p. We have to give a criterion for ρ to arise from a newform $f \in S_k(\Gamma_1(Np^2))$ such that $\pi_p(f)$ is principal series, for all weights $k \geqslant 2$. Just as in the proof of Theorem 1.2 above, the main step is to give a criterion for irreducible mod p Galois representations to arise from a newform $g \in S_k(\Gamma_1(Np))$, with $\pi_p(g)$ principal series. Here the essential observation is that one may reduce to the weight 2 case by means of a group cohomology argument, which is quite similar to the constructions carried ou by Hida [H1].

We study group cohomology for this. We have to consider the cohomology groups $H^1(\Gamma_1(N) \cap \Gamma_0(p), \mathbb{L}^n(\bar{\mathbb{F}}_p) \otimes \psi)$ where ψ is some power of the mod p

https://doi.org/10.1023/A:1000424731365 Published online by Cambridge University Press

cyclotomic character χ and $n = k - 2$. Here $\mathbb{L}^n(\bar{\mathbb{F}}_p)$ is the n-th symmetric power of the standard 2-dimensional representation over $\bar{\mathbb{F}}_p$ of $\Gamma_1(N) \cap \Gamma_0(p)$, and $\mathbb{L}^n(\bar{\mathbb{F}}_p) \otimes \psi$ denotes that the action on the module $\mathbb{L}^n(\bar{\mathbb{F}}_p)$ (which we will abbreviate below to $\mathbb{L}^n$) has been twisted by the character ψ of $\Gamma_1(N) \cap \Gamma_0(p)$, sending a matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

to $\psi(d)$. We consider ψ as a character on $(\mathbb{Z}/p)^*$. We identify, in the standard way (cf. [H]), $\mathbb{L}^n(\bar{\mathbb{F}}_p)$ with the homogeneous polynomials of degree n, with coefficients in $\bar{\mathbb{F}}_p$, in the variables X, Y.

Now we notice that, as a $\Gamma_1(N) \cap \Gamma_0(p)$ module, $\mathbb{L}^n \otimes \psi$ is reducible. More precisely, the subspaces of the form $\mathbb{L}_i^n := Y^i \mathbb{L}^{n-i}$ are left stable by $\Gamma_1(N) \cap \Gamma_0(p)$, with the successive quotients $\mathbb{L}_i^n / \mathbb{L}_{i+1}^n$ isomorphic as $\Gamma_1(N) \cap \Gamma_0(p)$ modules to $\bar{\mathbb{F}}_p(\psi\chi^{n-2i})$, where i ranges from $i = 0, 1, \ldots, n$, and with χ the mod p cyclotomic character. The short exact sequence

$$0 \to Y^{n-i-1}\mathbb{L}_{i+1}^n \to Y^{n-i}\mathbb{L}_i^n \to Y^{n-i}\mathbb{L}_i^n / Y^{n-i-1}\mathbb{L}_{i+1}^n \to 0,$$

of $\Gamma_1(N) \cap \Gamma_0(p)$ modules, gives a long exact sequence of cohomology, the part relevant to us being:

$$H^0(\Gamma_1(N) \cap \Gamma_0(p), \bar{\mathbb{F}}_p(\psi\chi^{n-2i})) \to H^1(\Gamma_1(N) \cap \Gamma_0(p), Y^{n-i-1}\mathbb{L}_{i+1}^n \otimes \psi)$$
$$\to H^1(\Gamma_1(N) \cap \Gamma_0(p), Y^{n-1}\mathbb{L}_i^n \otimes \psi) \to H^1(\Gamma_1(N) \cap \Gamma_0(p), \bar{\mathbb{F}}_p(\psi\chi^{n-2i}))$$
$$\to H^2(\Gamma_1(N) \cap \Gamma_0(p), Y^{n-i-1}\mathbb{L}_{i+1}^n \otimes \psi).$$

We need to study the above exact sequence with respect to the Hecke action on group cohomology (see, for instance, Section 3 of [H1] or Section 6.3 of [H2], for the definition of Hecke action on group cohomology). The map

$$\alpha\colon H^1(\Gamma_1(N) \cap \Gamma_0(p), Y^{n-i}\mathbb{L}_i^n \otimes \psi) \to H^1(\Gamma_1(N) \cap \Gamma_0(p), \bar{\mathbb{F}}_p(\psi\chi^{n-2i}))$$

is not equivariant for the Hecke action, but we claim that $\alpha T_q = q^i T_q \alpha$ where q is a prime, coprime to Np, and the operator T_q on the left hand side (resp. on the right hand side) is the action of the q-th Hecke operator on $H^1(\Gamma_1(N) \cap \Gamma_0(p), Y^{n-i}\mathbb{L}_i^n \otimes \psi)$ (resp., $H^1(\Gamma_1(N) \cap \Gamma_0(p), \bar{\mathbb{F}}_p(\psi\chi^{n-2i}))$).

To verify the claim, we compute explicitly the Hecke action on the cohomology groups. For convenience we denote $\Gamma_1(N) \cap \Gamma_0(p)$ by Γ. The action of T_q arises from the double coset $\Gamma\gamma^q\Gamma$ where γ^q is the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & q \end{pmatrix}.$$

We can write this double coset as the disjoint union of the coset $\Gamma\gamma_j (j = 0, \ldots, q - 1)$ and the coset $\Gamma\gamma_q$, where $\gamma_j (j = 0, \ldots, q-1)$ is the matrix

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

and γ_q is the matrix

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

with $\delta_q \in SL_2(\mathbb{Z})$ congruent to

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

mod N, and the identity matrix mod p. For $\tau \in \Gamma$, and $0 \leqslant j \leqslant q$, we have $\gamma_j\tau = \tau_j\gamma_\ell$, for some τ_j in Γ, and for some $\ell, 0 \leqslant \ell \leqslant n$. Then the action of T_q on a cocycle $u \in H^1(\Gamma_1(N) \cap \Gamma_0(p), Y^{n-i}\mathbb{L}_i^n \otimes \psi)$ is given by $u|T_q(\tau) = \sum_0^q \gamma_j^\iota u(\tau_j)$, where ι denotes the main involution (see Section 6.3 of [H2]). By the definition of α, we see that

$$\begin{aligned}\alpha(u|T_q(\tau)) &= \sum_0^{q-1} q^i\alpha(u(\tau_j)) + \psi(q)q^{n-i}\alpha(u(\tau_q)) \\ &= q^i\left(\sum_0^{q-1} \alpha(u(\tau_j)) + \psi(q)q^{n-2i}\alpha(u(\tau_q))\right).\end{aligned}$$

But the latter in turn, by the definition of the Hecke action on $H^1(\Gamma_1(N) \cap \Gamma_0(p), \bar{\mathbb{F}}_p(\psi\chi^{n-2i}))$ (see Section 3 of [H1]), is equal to $q^i\alpha(u)|T_q(\tau)$. This proves the claim.

Note the standard fact that H^0 and H^2 of the group cohomology are *Eisenstein* as Hecke modules. By this we mean that any maximal ideal $\mathfrak{m}$ of the Hecke algebra $\mathbb{T}_i$, where i is either 0 or 2, acting on H^0 and H^2 resp. of the group cohomology with suitable coefficients as above, is such that the corresponding mod p Galois representation

$$\rho_{\mathfrak{m}}: G_{\mathbb{Q}} \to GL_2(\mathbb{T}/\mathfrak{m})$$

associated to it, is reducible. Such maximal ideals $\mathfrak{m}$ are called Eisenstein. Here $\mathbb{T}_i$ is defined to be the $\bar{\mathbb{F}}_p$-subalgebra in the ring of endomorphisms of the corresponding cohomology group, generated by the action of the Hecke operators T_q (defined analogously to the action on H^1 above). The fact that H^0 and H^2 are Eisenstein can be deduced, for instance, from Section 1.8 and Section 1.9 of [H].

We consider the Hecke algebra generated by T_q for primes q with $(q, Np) = 1$. Thus, by the long exact sequence of cohomology above, and our computation of the Hecke action, we see that a non-Eisenstein maximal ideal $\mathfrak{m}$ of the Hecke algebra acting on $H^1(\Gamma_1(N) \cap \Gamma_0(p), \mathbb{L}^n(\bar{\mathbb{F}}_p) \otimes \psi)$ is in the support of one of $H^1(\Gamma_1(N) \cap \Gamma_0(p), \bar{\mathbb{F}}_p(\psi\chi^{n-2i})) \otimes \chi^i (0 \leqslant i \leqslant n)$, and conversely. Here by $\otimes\chi^i$ we mean that the Hecke action has been twisted by the i-th power of the determinant character.

This yields the following theorem.

THEOREM 1.3. *An irreducible mod $\mathfrak{p}$ representation arises from $f \in S_k(\Gamma_1(N) \cap \Gamma_0(p), \varepsilon)$, with ε a character of conductor p and $\pi_p(f)$ principal series, if and only if $\rho \otimes \omega^{-i}$ arises from $S_2(\Gamma_1(Np))$ for some i such that* $0 \leqslant i \leqslant k-2$.

Proof. The proof follows from the above discussion on noting that a newform $f \in S_2(\Gamma_1(N) \cap \Gamma_0(p), \varepsilon)$ has the property that $\pi_p(f)$ is principal series.

Note. The referee has remarked that the sufficiency can also be proved using multiplication by Eisenstein series.

By this theorem, Propositions 8.13 and 8.18 of [G], and [E], we have a complete determination of which mod p irreducible representations arise from $S_k(\Gamma_1(N) \cap \Gamma_0(p), \psi)$, for a given character ψ of conductor dividing p, at least for ψ a non-trivial character.

For unramified principal series lifts of weight k we have to look at $H^1(\Gamma_1(N), \mathbb{L}^n(\mathbb{F}_p))$, as in this case we only want to look at f in $S_k(\Gamma_1(N))$ which gives rise to ρ. We may write down the irreducible components in the Jordan-Holder series for $\mathbb{L}^n(\mathbb{F}_p)$ by using Lemma 3.2 of [AS]. Then using the proof of the weight part of Serre's conjecture in [E], we may give a criterion for ρ to arise from $S_k(\Gamma_1(N))$. This will be justified in the sequel to this paper.

After this, by the same method we employed for proving Theorem 1.2, we can determine the types $\pi_p(f) = \pi(\xi_1, \xi_2)$, where $\xi_1|\mathbb{Z}_p^* = \omega^a\varepsilon$ and $\xi_2 = \omega^d\varepsilon$ for character ε of p-power conductor and p-power order, for newforms f of weight k which give rise to ρ.

REMARK 1.4. We point out in passing that Carayol's lemma is not valid in general for p-new quotients of spaces of cusp forms of certain levels (this is not surprising in view of the existence of congruences between p-old and p-new forms!). This may be seen concretely from the fact than an irreducible modular mod p representation ρ, with $k(\rho) = p+1$, arises from the p-new part of $S_2(\Gamma_1(Np^2))$, for some integer N prime to p, but it does not arise from the p-new part of $S_2(\Gamma_1(N) \cap \Gamma_0(p^2))$ for such a N. This follows from Theorem 3 of [K].

2. Special Liftings

THEOREM 2.1. *Suppose that ε is a character of $\mathbb{Z}_p^*$ of conductor p^r with $r \geqslant 0$, and write $\varepsilon \equiv \chi^a$ (mod $\mathfrak{p}$) for some integer a. Let ρ be an irreducible* mod p *representation arising from $S_k(\Gamma_1(N))$ from some weight k. The a necessary and sufficient condition for there to exist a newform f of weight 2 and level dividing Np^{2r}, such that $\pi_p(f)$ is isomorphic to $\varepsilon \otimes$ sp and which gives rise to ρ, is that $\rho|D_p$ be of the form:*

$$\begin{pmatrix} \chi^{a+1}\psi & * \\ 0 & \chi^a\psi \end{pmatrix},$$

for some unramified character ψ of D_p.

Proof. The proof follows upon using results of [E], the theorem of Deligne (Theorem 2.5 of [E]), and a slight, and similarly proven, variant of part 1 of Theorem3 of [K] (this part of that theorem is due to Ribet). This variant would say that if a ρ, as above arises from a newform $h \in S_2(\Gamma_1(N))$ (with nebentype ε_h), then it also arises from the p-new part of $S_2(\Gamma_1(N) \cap \Gamma_0(p))$ iff $a_p(h)^2 \equiv \varepsilon_h(p)$ (mod $\mathfrak{p}$), where ε_h is the nebentypus of h.

We give the details. The necessity of the condition in the theorem follows from the fact that if ρ arises from a p-new form in $S_2(\Gamma_1(N) \cap \Gamma_0(p))$, then ρ also arises from $g \in S_{p+1}(\Gamma_1(N))$ by Proposition 8.18 of [G]. Further the p-th coefficient is forced to be a unit at the valuation given by $\mathfrak{p}$. Then the theorem of Deligne forces $\rho|_{D_p}$ to be of the form:

$$\begin{pmatrix} \chi\psi' & * \\ 0 & \psi \end{pmatrix},$$

for some unramified character ψ, ψ' of D_p. One further knows that $\psi(p)$ gives the p-th Fourier coefficient of g. We quote the classical result of Hecke that a p-new newform h of $S_2(\Gamma_1(N) \cap \Gamma_0(p))$ is such that $a_p(h)^2 = \varepsilon_h(p)$, with notation as in the preceeding paragraph. Thus we get that $\psi = \psi'$.

For the sufficiency, we use the fact that, assuming that $\rho|_{D_p}$ is of the form in the theorem, then [E] together with Proposition 8.18 of [G], gives that $\rho \otimes \chi^{-a}$ arises from a newform $h \in S_2(\Gamma_1(N) \cap \Gamma_0(p))$. Further, from the shape of $\rho|_{D_p}$, we see that $a_p(h)^2 \equiv \varepsilon_h(p)$ (mod $\mathfrak{p}$) where ε_h is the nebentypus of h. Then the result of Ribet, which is explained in part 1 of Theorem 3 of [K], yields that fact that $\rho \otimes \chi^{-a}$ also arises from a p-new form in $S_2(\Gamma_1(N) \cap \Gamma_0(p))$. From this the sufficiency follows.

3. An Example

The purpose of this section is to discuss an example to show that, what may be regarded as an implication in the opposite direction to that of Carayol's lemma, is false. To construct this example we use some of the results of [K]. This aslo shows that congruences in the (p, p) case behave differently, in that it is difficult to guess *a priori* the nature of the $\pi_p(f)$'s of newforms f which give rise to ρ (but see the forthcoming sequel to this paper).

Thus we start from a newform f in $S_2(\Gamma_0(pq))$, which is pq-new where q is a prime distinct from p. Denote the automorphic representation corresponding to f by π_f. We assume that the mod p representation ρ attached to f is irreducible, and further that it is finite at p, but such that $\rho|I_p$ is not semisimple. Newforms f which meet these requirements may be constructed by starting from a newform h in $S_2(\Gamma_0(q))$ with the corresponding mod p representation ρ irreducible, such that $\rho|I_p$ is not semisimple, and $a_p(h) \equiv \pm 1$ (mod $\mathfrak{p}$) for the fixed place $\mathfrak{p}$ above p. Then we see by [R] that h is congruent (in the sense of the Fourier coefficients at almost all primes being congruent modulo $\mathfrak{p}$) to a newform of the required kind in $S_2(\Gamma_0(pq))$). By [R1] it follows that this mod p representation cannot be finite at q. Denote the indefinite quaternion algebra ramified exactly at p and q by B, and denote $\bar{G}$ the algebraic group over $\mathbb{Q}$ defined by the multiplicative group B^*. Denote by $\bar{\pi}_f$, the automorphic representation of the group given by the adelic points of $\bar{G}$ that corresponds to π_f by the Jacquet–Langlands correspondence.

With our assumptions on f, the main theorem of [K], see the introduction of loc. cit., yields that the mod p representation attached to f, arises from a newform g in the p-new part of $S_2(\Gamma_0(qp^2))$. This form g is q-new by what we have noted above that the mod p representation corresponding to f is ramified at q. We claim that $\pi_p(g)$ is supercuspidal. This easily follows from Theorem 1.2 above. Namely, $\pi_p(g)$ cannot be principal series by the classification in Theorem 1.2, and one can rule out the case that $\pi_p(g)$ is twist of special by inspection.

Denote by $\bar{\pi}_g$ the automorphic representation of the group given by the adelic point of $\bar{G}$ that corresponds to π_g by the Jacquet–Langlands correspondence (as g is discrete series at p and q). From what we have said so far we conclude, as in Section 5.3 of [C], (in this results of Section 5 of [Ge] get used) that

$$\bar{\pi}_p(g) = \mathrm{Ind}(B_p^*, \Omega\bar{K}_p^1, \bar{\xi}'),$$

with notation exactly as in Section 5.3 of loc. cit., which we reprise below. Thus $B_p = B \otimes \mathbb{Q}_p, \mathcal{O}_p$ is the ring of integers of $B_p, \mathcal{M}_p$ is the maximal ideal of this ring, $\bar{K}_p^0$ is the group of units of $\mathcal{O}_p, \bar{K}_p^1 := 1 + \mathcal{M}_p, \Omega$ is the quadratic unramified extension of $\mathbb{Q}_p$ in a fixed algebraic closure, and $\bar{\xi}'$ is a character of conductor p of Ω^* (which does not factor through the norm), which may also naturally regard as a character of $\Omega\bar{K}_p^1$. Thus the order of $\bar{\xi}'$ restricted to $\bar{K}_p^0$ divides $p^2 - 1$, and is thus prime to p, which will be crucial for us. As noted in Section 5.3 of loc. cit., $\bar{\pi}_p(g)$ is then an irreducible representation of dimension 2, whose restriction to $\bar{K}_p^1$ is trivial, and $\bar{\xi}'$ may be regarded as a character of $\bar{K}_p^0/\bar{K}_p^1$, which may be identified with $\mathbb{F}_{p^2}^*$. Then the above representation is decomposed as a direct sum of $\bar{\xi}^0$ and $\bar{\xi}^{0^\sigma}$ (σ is the non-trivial automorphism of $\mathbb{F}_{p^2}$) as a representation of $\mathbb{F}_{p^2}^*$, where the superscript 0 just denotes that we are considering the character which $\bar{\xi}'$ naturally induces on $\mathbb{F}_{p^2}^*$. We remark that as $\bar{\xi}^0$ is of order prime to p, its reduction mod p, which we denote by ξ, is non-trivial. In fact ξ also does not factor through the norm character, as is easily checked.

We consider $\bar{K}^1$, a compact open subgroup of $\bar{G}(\mathbb{A}_f)$, where we take $\bar{K}_v^1$ to be a maximal open compact subgroup at all places other than p (we as usual fix isomorphisms $\bar{G}(\mathbb{Q}_v) = GL_2(\mathbb{Q}_v)$ for v other than p or q etc.) and to be $\bar{K}_p^1$ at p. Just as in Section 5.4 of loc. cit. we denote the Shimura curve corresponding to $\bar{K}^1$ by S^1 (though note that our $\bar{K}^1$ is slightly different). This curve has the natural action of $\bar{\Gamma} := \bar{K}_p^0/\bar{K}_p^1 (= \mathbb{F}_{p^2}^*)$.

As in Section 5.6 of [C], we see that the p-adic Galois representations corresponding to $\bar{\pi}_f$ and $\bar{\pi}_g$ intervene in the étale cohomology group $H^1_{\text{ét}}(S^1_{\bar{\mathbb{Q}}}, \bar{\mathbb{Q}}_p)$, such that the corresponding mod p Galois representations, which we denote by ρ_f and ρ_g, intervening in $H^1_{\text{ét}}(S^1_{\bar{\mathbb{Q}}}, \bar{\mathbb{F}}_p)$ are isomorphic. But from what we have noted about the nature of $\bar{\pi}_p(f)$ and $\bar{\pi}_p(g)$, we claim that the induced action of $\bar{\Gamma}$ is trivial on ρ_f (as $\bar{\pi}_p(f)$ is one-dimensional), while we may choose the copy ρ_g of ρ in $H^1_{\text{ét}}(S^1_{\bar{\mathbb{Q}}}, \bar{\mathbb{F}}_p)$ so that $\bar{\Gamma}$ acts by the character ξ on ρ_g. (We have already noted that ξ is a non-trivial character, and more, in that it does not even factor through the norm character.)

To see this, for instance the fact that ρ_g can be chosen so that the action of $\bar{\Gamma}$ on it is by the character ξ, follows from the comparison theorem between étale and ordinary cohomology of the curve S^1. This comparison isomorphism is compatible with the induced $\bar{\Gamma}$ and Hecke action on these cohomologies. Further, g gives rise to a non-zero element in $H^1(S^1_{\mathbb{C}}, \mathbb{C})$ which is an eigenvector for the Hecke action on this cohomology group, with the same eigenvalues as g, and on which $\bar{\Gamma}$ acts by the character $\bar{\xi}'$. From this the claim follows. (We thank the referee for pointing out a mistake in this part of an earlier version.)

Thus, in conclusion, we see that we have constructed two isomorphic, irreducible 2-dimensional mod p representations in the étale cohomology of a Shimura curve, which carries a $\mathbb{Q}$-rational action of a finite abelian group, such that the action of this abelian group on these two isomorphic representations is via different (non-conjugate) characters.

REMARK 3.1. In fact we may further ask for a classification of the characters ξ by which $\bar{\Gamma}$ can act on copies of ρ in $H^1_{\text{ét}}(S^1_{\bar{\mathbb{Q}}}, \bar{\mathbb{F}}_p)$. This, in principle, can be answered by the method that we have briefly outlined in the introduction, and which will be presented in detail in a sequel.

4. Remarks on our Earlier Paper

In this section we sharpen the main theorem of [K] (stated in the introduction of loc. cit.).

There is a question about congruences in the (p,p) case which was avoided in the study in [K], as represented in the main theorem of the paper stated in the introduction of loc. cit. There it was stated that if f is a newform in $S_2(\Gamma_1(Np))$ which give rise to an irreducible mod $\mathfrak{p}$ representation ρ, such that $2 < k(\rho) < p+1$, then some Galois conjugate of ρ arises a p-new form in $S_2(\Gamma_1(Np) \cap \Gamma_0(p^2))$. In the case when ρ is ordinary, what has been said in [K] is not enough to yield more precise information, i.e., one cannot delete the phrase *some Galois conjugate of*. We now fill in this lacuna using the methods of [K].

THEOREM 4.1. *Let f be a newform in $S_2(\Gamma_1(Np))$ and assume that the* mod *p representation ρ (with respect to the chosen place $\mathfrak{p}$ above p) attached to f is irreducible and is such that $2 < k(\rho) < p+1$. Then ρ also arises from the p-new part of $S_2(\Gamma_1(Np) \cap \Gamma_0(p^2))$.*

Proof. We use the technique of the proof of Theorem 2 in [K], the original idea of which goes back to Ribet's ICM paper, cf. [R]. Thus we consider the natural degeneracy map $\alpha\colon J_1(Np)^2 \to J(Np, Np^2)$. The latter is the Jacobian of the complete curve made from $\Gamma_1(Np) \cap \Gamma_0(p^2)$. The composite $\alpha' \cdot \alpha$, where α' is the dual to α, is important for us. When viewed as a 2×2 matrix of endomorphisms of $J_1(Np)^2[p]$, it is given by:

$$\alpha' \cdot \alpha = \begin{pmatrix} 0 & U_p \\ U_p' & 0 \end{pmatrix}.$$

This follows from a standard computation (here U_p') is the dual of U_p). Now, because $k(\rho)$ is strictly between 2 and $p+1$, the representation ρ only occurs in A, the abelian subvariety of $J_1(Np)$ whose cotangent space is spanned by forms with non-trivial nebentype at p. On A we have the relation $U_pU_p' = p$, as is well-known. Thus a Hecke stable Galois module V of $J_1(Np)$, which is isomorphic to ρ, is annihilated either by U_p or U_p'. We suppose that it is annihilated by U_p', the other case being entirely similar. Then $(V, 0)$ is in the kernel of the above matrix. As V is by assumption irreducible we see, by the same method as in the proof of Theorem 2 of [K], that ρ intervenes in the intersection of the p-old and p-new subvarieties of $J(Np, Np^2)$. For this we will have to use a variant of Theorem 1 of [K] (similarly proven) to the effect that the group of connected components of the natural degenerancy map $J_1(Np)^2 \to J(Np, Np^2)$ is Eisenstein as a Hecke module. From this the theorem follows.

CORRECTION 4.2. As pointed out by the referee of the present paper, in the proof of Theorem 2 of [K], we have transposed T_p and T_p^* in the formula for the matrix R. Fortunately this does not affect the proof there materially, as T_p and T_p^* play 'symmetric roles', and the proof can be easily corrected (for instance, instead of considering $(V, 0)$, we have to consider $(0, V)$ in the proof of loc. cit.).

Acknowledgements

The author is grateful to Dipendra Prasad for helpful conversations and encouragement. He also thanks MRI for hospitality during work on this paper. Thanks are also due to the referee for his careful reading of the manuscript and pointing out its inadequacies.

References

[AS] Ash, A. and Stevens, G.: Modular forms in characteristic ℓ and special values of L-functions, *Duke Math. J.* (1986), 849–868.

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

[D] Diamond, F.: The refined conjecture of Serre, *Proc. of Conf. on Elliptic Curves and Modular Forms*, Hong Kong 1993, International Press.

[DT] Diamond, F. and Taylor, R.: Lifting modular mod ℓ representations, *Duke Math. J.* 74 (1994), 253–269.

[E] Edixhoven, B.: The weight in Serre's conjectures on modular forms, *Inv Math.* 109 (1992), 563–594.

[G] Gross, B.: A tameness criterion for Galois representations associated to modular forms (mod p), *Duke Math. J.* 61 (1990), 445–517.

[Ge] Gérardin, P.: Facteurs locaux des algèbres simples de rang 4, I, in *Groupes Réductifs et Formes Automorphes*, *Publ. Math.* Paris VII (1978), 37–77.

[H] Hida, H.: Congruences of cusp forms and of the special values of their zeta functions, *Inv. Math. 63* (1981), 225–261.

[H1] Hida, H.: On congruence divisors of cusp forms as factors of the special values of their zeta functions, *Inv. Math.* 64 (1981), 221–262.

[H2] Hida, H.: Elementary theory of L-functions and Eisenstein series, *LMSST* 26, Cambridge University Press, 1993.

[K] Khare, C.: Congruences between cusp forms: the (p, p) case, *Duke Math. J.* 80 (1995), 31–68.

[R] Ribet, K.: Congruence relations between modular forms, *Proc. ICM*, Warsaw (1983), 503–514.

[R1] Ribet, K.: Report on mod ℓ representations of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, in Ui. Jannsen et al. (eds.), *Motives*, Part 2 (1994), 639–676.

[S] Serre, J-P.: Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, *Duke Math. J.* 54 (1987), 179–230.