Congruences between Cusp Forms

Thesis by
Chandrashekhar B. Khare

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

California Institute of Technology
Pasadena, California
1995

(Submitted January 26, 1995)

Acknowledgement

It is my pleasure to acknowledge the debt of gratitude I owe to many people. I would like to thank Prof. Hida for his constant guidance and patience. I would like to thank Dipendra Prasad for his encouragement, for helpful correspondence and discussions and for allowing me to include our joint work in Chapter 3. I would like to thank Prof. D. Ramakrishnan for his help and support. I would like to thank my family for its unstinting support.

Contents

Abstract

In this thesis we study the ring of modular deformations of an absolutely irreducible mod p representation which is modular by studying the congruences between newforms of weight 2 and varying p power levels. This fills in a missing case in the literature of the study of congruences between modular forms of varying levels. The results of §1.3, §1.4 and §1.5 give a thorough analysis of congruences in the (p, p) case. The results we prove along the way in Chapter 1 shed light on the multiplicities with which certain 2 dimensional representations arise in the Jacobians of modular curves. In §1.7 we apply the study of congruences in the (p, p) case to prove lower bounds on the ring of modular deformations. This lower bound has been proven earlier in Gouvea.

In Chapter 2 we study local components of Hecke algebras which arise by studying Hecke action on the space of mod p modular forms of fixed level and all weights. We relate the computation of dimensions of ring of modular deformations to certain properties of Hecke exact sequences. These exact sequences arise from the phenomenon that mod p there are inclusions between modular forms (identified with their q–expansions) of different weights.

In Chapter 3, which is joint work with D. Prasad, we raise a natural question about the nature of Fourier coefficients of cuspidal eigenforms which may be viewed as asking for a version of the Chinese Remainder Theorem for automorphic representations and answer it in some simple cases.

Chapter 0

Introduction

It has been known for some time, as a consequence of the work of numerous mathematicians, that newforms for congruence subgroups of $SL_2(\mathbb{Z})$ give rise to a compatible system of ℓ-adic representations, and if the p-adic representations attached to two newforms are isomorphic for any prime p, then the newforms are in fact equal. But the corresponding statement is not true for the mod p reductions of p-adic representations attached to newforms, as different newforms can give rise to isomorphic mod p representations which arise from reduction mod p of the corresponding p adic representations (this is well defined if we assume that the mod p representation is absolutely irreducible). This is a reflection of the fact that distinct newforms can be congruent modulo p. To study the different levels from which a given modular mod p representation can arise is interesting, and has been much studied.

Thus if we consider the image of the classical Hecke operators in the ring of endomorphisms of the Jacobian $J_0(S)$ of the modular curve $X_0(S)$ then the resulting $\mathbb{Z}$ algebra is of finite rank over $\mathbb{Z}$. We denote it by $\mathbb{T}_S$. Then to any maximal ideal m of $\mathbb{T}_S$ of residue characteristic say p, we may attach, after the work of Eichler-Shimura, a representation:

$$\rho_m : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL_2(\overline{\mathbb{T}_S/m})$$

(we shall assume that ρ_m is absolutely irreducible) such that it is unramified at all primes r prime to pS, and for such primes $\mathrm{tr}(\rho_m(\mathrm{Frob}_r))$ is the image of T_r in $\mathbb{T}_S/m$ and $\det(\rho_m(\mathrm{Frob}_r)) = r$. Then on viewing ρ_m abstractly, one may try to classify all the pairs $(\mathbb{T}_M, n)$, where n is a maximal ideal of $\mathbb{T}_M$, which give rise (in the above fashion) to a representation isomorphic to ρ_m in a non trivial way (i.e., n should be associated to a newform of level M). This classification has been essentially carried

out in the work of several people – Mazur, Ribet, Carayol, Diamond, Taylor – for all M prime to p. In Chapter 1 of this thesis we study the case when we do not impose this condition. We shall colloquially talk of this as the (p,p) case.

This case differs in many salient points. It follows from the classification of Carayol that the exponent with which any prime ℓ different from p occurs in the factorisation of any M as above is bounded. As a consequence of the more precise result we prove in Chapter 1 (this is the heart of the thesis), we see that arbitrarily large powers of p can divide such a M. We present a complete analysis of congruences in the (p,p) case in this chapter. The (p,p) case is anomalous in various ways. For instance one cannot use the theorem of Langlands and Carayol to guess the congruences which arise in this case as one may do in the non (p,p) case. Because of Theorem 2 and Theorem 3 of Chapter 1, we know more about this. Theorem 2 essentially says that levels can always be raised to levels with high powers of p in them while Theorem 3 analyses the case of low p powers.

The determination of the levels M from which ρ_m can arise as above is essentially a question of studying congruences mod p between forms of varying levels. This has been studied mostly after the original method of Ribet. According to this, one has to study in detail the degeneracy maps and in the most crucial step, determine the kernel of a natural degeneracy map $J_0(M)^2 \to J_0(Mp)$. The (p,p) case is anomalous here too, because if M is prime to p this kernel is finite and in a certain technical sense irrelevant, but if M is not prime to p this kernel in general contains an abelian variety. In Theorem 1 of Chapter 1 we pin this kernel down using the theory of modular symbols. This allows us to prove Theorem 1. We believe that the use of modular symbols in this context, which has been initiated in this thesis, will prove to have more applications. We also analyse in §1.4 of this chapter the minimal p power level which divides any level M which can give rise to ρ_m.

The (p,p) case is different in yet another way. As, it is known that if m is a maximal ideal of $\mathbb{T}_S$, and the residue characteristic m is prime to $2S$, then $J_0(S)[m]$ is of dimension 2 over $\mathbb{T}_S/m$. But we prove that in the case when the residue characteristic of m is p and m is non ordinary at p, then the dimension of

$J_0(S)[m]$ over $\mathbb{T}_S/m$ tends to infinity as the power of p which divides S tends to infinity.

We also apply our study of congruences between forms of weight 2 in the (p,p) case to study the ring of modular deformations of ρ_m. There is an elaborate theory of this due to Hida in the case when one is studying ordinary modular deformations. In the recent work of Wiles the Galois deformations of ρ_m with certain properties are shown to be modular. But one of the motivating problems of this thesis was to study, in Mazur's original universal deformation ring, the locus corresponding to modular deformations. In this case one does not expect all deformations to be modular and some work of Hida [H 2] strongly suggests that in fact the Krull dimensions of Mazur's ring and that of the ring of modular deformations are different. We have not succeeded in proving this but apply our results in the last section of Chapter 1 to give a lower bound on the ring of modular deformations, which is probably also the right dimension of the ring of modular deformations. This lower bound has been proven earlier in [G].

In order to study more closely this dimension, in Chapter 2 we study mod p modular forms of fixed level and all weights. We strengthen some results of Jochnowitz about local components of Hecke algebras, and relate the question of determining the structure of local components, and in particular their dimension, to studying certain naturally occurring extensions of Hecke modules.

In Chapter 3 (which is joint work with D. Prasad) we study another aspect of eigen cuspforms. The Fourier coefficients of cuspidal Hecke eigenforms are rather mysterious and little is known about them. We pose a converse question to Deligne's theorem which proves bounds on these eigenvalues. This can be viewed as asking for a version of the Chinese Remainder Theorem in the context of automorphic representations. We answer this in the only case in which there is a known way to explicitly construct eigen cuspforms, i.e., in the CM case. We also pose and answer in simple cases a similar question about Galois representations.

Chapter 1

Congruences between cusp forms: the (p,p) case

1.1. Introduction

The question of raising levels of a cuspidal eigenform for a congruence subgroup of $SL_2(\mathbb{Z})$ has been studied in many papers after the method inaugurated by Ribet [R]. Raising of levels in various contexts has found many applications: for instance, in the lowering of levels, cf. [R 1], in attaching Galois representations to Hilbert modular forms, cf. [T]. Perhaps motivated by these applications, the question of raising levels has mainly been studied for raising the level by p modulo a prime ℓ, when p and ℓ are distinct. In this chapter we would like to study the so called (p,p) case, which in a certain sense completes the picture which has emerged in the papers [R], [D-T]. We state this theorem below. We may view this as saying that cuspforms of varying p power level are tightly webbed together by many congruences mod p in a systematic way. Our method of proof follows the procedure for raising levels in [R], but differs in one significant detail. At one of the crucial points in the proof in [R], a lemma of Ihara was used to control the error terms and show them to be irrelevant, i.e., Eisenstein. We do not have recourse to this lemma at a similar point in our proof of the theorem below. So we have to take a different tack and use instead the modular symbols isomorphism of Manin, as interpreted in the paper of Ash and Stevens [A-S]. The use of the modular symbols isomorphism performs the task of converting a problem in degree 1 cohomology to what turns out to be an easier problem which belongs to degree 0 cohomology of a different situation. This is similar to the use of the Jacquet-Langlands correspondence in [D-T], which allows one in certain situations of congruences which are studied there, to switch to a definite quaternion algebra where controlling the error terms becomes easier for a similar reason. Our study of this (p,p) case was motivated by the effort to understand the ring of modular deformations of a mod p representation which is

absolutely irreducible and is modular, i.e., arises in the well known way from a maximal ideal of a Hecke algebra. We will give some applications of our theorem to this situation. It is perhaps time to state the theorem.

Theorem. *Suppose f is a newform of weight 2 for the group $\Gamma_0(Np^r)$, $(N,p)=1$ and that the mod p representation f gives rise to is absolutely irreducible. Then for any $s > max(r,2)$, there exists a newform g, of weight 2 for the group $\Gamma_0(N'p^s)$, where $N'|N$, such that:*

$$f \equiv g \pmod{\mathfrak{p}},$$

where $\mathfrak{p}$ is a place above p.

In the theorem, by a congruence between modular forms, we mean a congruence of the Fourier coefficients outside Np and in general throughout this chapter by a congruence mod p between newforms we shall mean that they give rise to isomorphic (absolutely irreducible) mod p representations.

We can then think of the question of lowering the p part of the level of absolutely irreducible mod p modular representations. The answer in principle should be a consequence of the proof in [E] of the weight part of Serre's conjectures as it is a well known principle that there is a correspondence between the minimal p part of the level of a newform of weight 2 mod p and the least weight at which it arises (mod p) from a level which is prime to p. But we make this explicit in §1.4. The above theorem still leaves open the question of mod p congruences between forms of level N, Np and Np^2. We study this in §1.5 and prove a theorem which completely settles this question (at least assuming that $p \geq 5$). Theorems 4 and 4′ analyse congruences in the (p,p) case further. The criterion for congruences between forms of low p power levels is a little involved and depends on finer properties of the corresponding mod p representation at p such as finiteness etc. Thus we see at the end of §1.5 that if we gather together all the results we have either proven or noted as existing in the literature then we have obtained a rather complete picture of congruences between newforms of weight 2 in the (p,p) case.

We now give a schematic outline of the chapter. In §1.2 we prove the main

technical result in the proof of the theorem above (which is also Thorem 2 of §1.3). We calculate the kernel of a natural degeneracy map:

$$J_0(Np^r)^2 \to J_0(Np^{r+1})$$

and show that (cf. Theorem 1 of §1.2) upto an error which we calculate to be Eisenstein, it is what one would expect, i.e., a copy of $J_0(Np^{r-1})$ considered as embedded in $J_0(Np^r)^2$ using degeneracy maps which we will make precise in §1.2. We note that a special case of this question has previously been studied by S. Ling in [L], which corresponds to putting $N = 1$ and $r = 1$. In §1.3 we apply the result in §1.2 to raise levels as in the above theorem. In §1.4 we analyse the question of lowering the p power level of a mod p representation (which is absolutely irreducible and modular) using the results of [E]. In §1.5 we prove a result about congruences between forms of low p power levels which corresponds to cases left out in Theorem 2 (cf. Theorem 3). This together with Theorems 2, 4, and 4′ and §1.4 gives a rather complete picture of congruences in the (p, p) case. In §1.6 we apply Theorem 2 (in a slightly generalised form, for which see Remark 5) to study the multiplicities with which certain two-dimensional representations occur in $J_0(Np^r)$. In §1.7 we use Theorem 2 to get information about the ring of modular deformations of an absolutely irreducible mod p representation which is modular.

1.2. Calculation of a kernel

We set up some notation. In what follows N is assumed to be prime to p and $r > 0$. We recall that there are the standard Atkin-Lehner degeneracy maps:

$$\alpha_1 : X_0(Np^{r+1}) \to X_0(Np^r)$$

and

$$\alpha_p : X_0(Np^{r+1}) \to X_0(Np^r).$$

These have the usual modular interpretation, i.e., on viewing $X_0(Np^{r+1})$ as associated to the (naive) moduli problem of classifying pairs $(E, C_{Np^{r+1}})$, where E is

an elliptic curve and $C_{Np^{r+1}}$ is a cyclic subgroup of E of order Np^{r+1}, and viewing $X_0(Np^r)$ analogously, α_1 is the map described by:

$$\alpha_1(E, C_{Np^{r+1}}) = (E, C_{Np^r})$$

and α_p is the map described by:

$$\alpha_p(E, C_{Np^{r+1}}) = (E/C_p, C_{Np^{r+1}}/C_p),$$

where C_{Np^r} and C_p are the cyclic subgroups of order Np^r and p respectively of $C_{Np^{r+1}}$.

By the Picard functoriality of the Jacobian, these degeneracy maps induce maps ${\alpha_1}^*$ and ${\alpha_p}^*$ from $J_0(Np^r)$ to $J_0(Np^{r+1})$, these being the Jacobians of the respective modular curves. These maps are injective as the covering $X_0(Np^{r+1}) \to X_0(Np^r)$ does not factor through any non-trivial unramified covering. We denote by α the sum of these two degeneracy maps. Thus α is the map:

$$\alpha : J_0(Np^r)^2 \to J_0(Np^{r+1})$$

which by definition is given by:

$$\alpha(x, y) = {\alpha_1}^*(x) + {\alpha_p}^*(y).$$

The image of α is called the p-old subvariety of $J_0(Np^{r+1})$, the quotient of $J_0(Np^{r+1})$ by the p-old subvariety is called its p-new quotient and the connected component of the kernel of the map dual to α which arises by the autoduality property of the Jacobian is called the p-new subvariety of $J_0(Np^{r+1})$. From the point of view of [R], to study congruences, amounts to understanding the canonical isogeny between the p-new quotient and the p-new subvariety.

Now we bring into play the Hecke action, our discussion being taken from §3 of [R 1], except that we use the symbols T_n for the Albanese action as we explain. For this it is convenient to start with a general modular curve $X_0(M)$ and its Jacobian $J_0(M)$. Shimura's ring $R(\Gamma_0(M), \Delta')$, cf. [S], induces self-correspondences of the

curve $X_0(M)$, which induce maps of $J_0(M)$ in two ways, one by using the Picard functoriality of the Jacobian and the other by using the Albanese functoriality. We shall consider the image of the Hecke ring, as the algebra of these correspondences is usually called, in $\mathrm{End}(J_0(M))$ using the action induced by the Albanese functoriality. We denote this ring by $\mathbb{T}_M$. It is a finitely generated $\mathbb{Z}$ algebra which is generated by the images of the correspondences T_n in the endomorphism ring, which arise from the sum of the correspondences induced by double cosets $\Gamma_0(M)\gamma\Gamma_0(M)$ where $\gamma \in \Delta'$, $det(\gamma) = n$, by Albanese functoriality. The endomorphism associated to this induced by the Picard action is denoted by ξ_n. These actions are related by the Rosati involution of the Jacobian and can be seen to be intertwined by the Atkin-Lehner involution w, i.e., $wT_nw = \xi_n$. If M divides M', and if they have the same radical, then there is a natural surjection from $\mathbb{T}_{M'}$ to $\mathbb{T}_M$, which takes T_n to T_n, obtained by restricting the action of the former T_n to the image of $J_0(M)$ in $J_0(M')$ under the natural degeneracy map. As $S_2(\Gamma_0(M))$ is identified with the space of regular differentials of $J_0(M)$ on viewing this as the Albanese variety of $X_0(M)$, T_n induces the classical operator T_n on $S_2(\Gamma_0(M)$ (this is the main reason that we depart from the convention of [R 1] and call T_n what is called ξ_n there). As an endomorphism of an Abelian variety is determined by its action on differentials, this association is one to one. The two actions of Hecke on the space of regular differentials are related by the main involution ι, i.e., if T_n corresponds to the double coset $\Gamma_0(M)\gamma\Gamma_0(M)$, then the action of ξ_n corresponds to the double coset $\Gamma_0(M)\gamma^\iota\Gamma_0(M)$.

Now we recall the correspondence between maximal ideals m of $\mathbb{T}_M$ and two-dimensional Galois representations over $\mathbb{T}_M/m$ occurring in $J_0(M)$. We assume that the Galois representation ρ_m associated by Eichler-Shimura to this, for more details see [R 1], is absolutely irreducible. We consider the finite dimensional $\mathbb{T}_M/m$ vector space $J_0(M)[m]$. Then as a $\mathbb{T}_M/m[\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})]$ module it is shown in [M 1] and [B-L-R], using the Eichler-Shimura relations, that $J_0(M)[m]$ is isomorphic to the direct sum of a certain number of copies, say q, of a two-dimensional vector space over $\mathbb{T}_M/m$ which as a Galois module is isomorphic to ρ_m. q is referred to as the mulitiplicity with which ρ_m occurs in $J_0(M)$. We signal an abuse of

notation that we will be guilty of in that we will refer to a maximal ideal of $\mathbb{T}_M$ and the corresponding one of $\mathbb{T}_{M'}$ which arises by pull back under the natural map $\mathbb{T}_{M'} \to \mathbb{T}_M$, with notation as above, by the same name.

We now revert to our specific situation of Jacobians of $X_0(Np^r)$, for varying $r \geq 1$. $\mathbb{T}_{Np^{r+1}}$ restricts to produce endomorphisms of the p-old and p-new subvarieties. We note that by our definition $\mathbb{T}_{Np^{r+1}}$ includes the pth Hecke operator which induces the operator U_p on cusp forms of weight 2 and with p dividing their level. Throughout this chapter we shall denote what is usually called U_p by T_p, as we are considering only forms at levels which are divisible by p. (As another general reference for the Hecke action we may refer to the discussion in Chapter 2 (§5.4 and §5.8) of [M-W].) We say that the action of Hecke on the p-new and p-old subvarieties factors through the p-new and p-old quotients respectively. For the moment we drop the subscript which denotes the level for which one considers the Hecke action. We call the p-new and p-old quotients $\mathbb{T}^{\,p-\text{new}}$ and $\mathbb{T}^{\,p-\text{old}}$ respectively. Then the phenomenon of congruences arises because the natural injection

$$\mathbb{T} \to \mathbb{T}^{\,p-\text{new}} \times \mathbb{T}^{\,p-\text{old}}$$

is not surjective (the image has finite index, the prime divisors of this index being those modulo which there will exist congruences between p-old and p-new forms). As the space of regular differentials of the p-old and p-new subvarieties are identified with the p-old and p-new forms respectively, this is sensible notation.

We denote by J the image of $J_0(Np^{r-1})$ in $J_0(Np^r)^2$ under the map β which is defined by:

$$\beta(x) = (-{\alpha_p}^*(x), {\alpha_1}^*(x)).$$

Here as in the above we are abusing notation by conflating the degeneracy maps arising from varying p power levels, as the gain in accuracy in putting subscripts of r all over the place may be outweighed by the resulting clutter of symbols. We note that the mapping β is injective and that J is in the kernel of α. Now we may state the main result of this section.

Theorem 1. *If we denote by A the kernel of α, the quotient A/J is a finite*

Hecke module which is Eisenstein, i.e., the group of connected components of A is Eisenstein.

Remark 1. By a module for the Hecke algebra being Eisenstein we mean that the maximal ideals which are in the support of the module give rise to Galois representations which are reducible. For the rest of this section, we will consider Hecke algebras which have been deprived of the Hecke operators corresponding to primes dividing the level for the technical reasons that the degeneracy maps we defined above are not equivariant with respect to the pth Hecke operator and that the action of the Hecke operators which involve primes dividing the level are not self-adjoint, i.e., the endomorphisms they induce on the Jacobian by Albanese and Picard functoriality need not be the same.

Proof. We begin by noting that J is the connected component of the identity of the kernel of α. This one may see by looking at the maps induced by the degeneracy maps on the space of regular differentials on $J_0(Np^r)$ viewing this as the Albanese variety of $X_0(Np^r)$. Then this space of differentials is identified with $S_2(\Gamma_0(Np^r))$. From this we conclude that J is the connected component of the identity by noting that if:

$$f(z) + g(pz) = 0$$

for $f, g \in S_2(\Gamma_0(Np^r))$ then $g \in S_2(\Gamma_0(Np^{r-1}))$ (cf. [A-L]), as then this tells us that J is the largest Abelian subvariety of the connected component of the identity of A. Thus we see that the quotient we are interested in studying, namely A/J, as a Hecke module, is identified with the group of connected components of A and as such is finite. We are interested in studying the exact sequence:

$$0 \to A \to J_0(Np^r)^2 \xrightarrow{\alpha} J_0(Np^{r+1})$$

and in calculating A. We instead study the analogous exact sequence in group cohomology:

$$0 \to K \to H^1(\Gamma_0(Np^r), M)^2 \xrightarrow{\alpha} H^1(\Gamma_0(Np^{r+1}), M). \tag{1}$$

Here we are again abusing notation and conflating the map α and its counterpart in group cohomology. We really have to study parabolic cohomology, but it is easily seen that it is enough to control the kernel K in (1) (K is defined by means of the exactness of (1)) and so we will work with ordinary cohomology. The above two sequences are well related if we are considering torsion free groups. But in the presence of torsion too $H^1(Y_\Gamma, M)$, with Y_Γ being the open curve associated to a congruence subgroup Γ of $SL_2(\mathbb{Z})$, is isomorphic to the subgroup of $H^1(\Gamma, M)$ (we shall only consider M with trivial Γ action) which consists of homomorphisms of $\overline{\Gamma}$ ($:= \Gamma/\pm 1$) which are trivial on the normal subgroup generated by the elliptic elements. Because of this it is easily seen that it is enough to study group cohomology to prove Theorem 1 even if the groups we are considering are not torsion-free. The underlying principle in what follows is that the kernels and cokernels of the natural maps between ordinary, compactly supported and parabolic cohomology have only maximal ideals of the Hecke algebra which are Eisenstein in their support. Here we consider the Hecke operators acting on cohomology groups via the standard action, cf. [S, Chapter 8]. So to study the localisation at a non-Eisenstein maximal ideal of any of these cohomology groups considered as Hecke modules, the particular kind of cohomology we study is not of consequence. Furthermore the modular symbols isomorphism provides a useful description of H^1 with compact support. We will now elaborate on this.

In (1) we are taking cohomology with respect to a module M on which the relevant group acts trivially. The module of interest to us will be $\mathbb{C}/\mathbb{Z}$. For the sake of being specific let us recall the degeneracy maps in the context of group cohomology. The analogs of ${\alpha_1}^*$ and ${\alpha_p}^*$, which we denote by the same symbols, are given by:

$$ {\alpha_1}^*(f)(x) = f(x) $$

$$ {\alpha_p}^* f(x) = f(\pi x \pi^{-1}) $$

where,

$$ f \epsilon H^1(\Gamma_0(Np^r), M),\ x\ \epsilon \Gamma_0(Np^{r+1}) $$

and the matrix π is:

$$\pi = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}.$$

We check easily that ${\alpha_1}^*(f)$ and ${\alpha_p}^*(f)$ are elements of $H^1(\Gamma_0(Np^{r+1}))$. Then we define α as before to be the sum ${\alpha_1}^* + {\alpha_p}^*$. We have an analogous map β which is defined as in §1.1 in terms of these degeneracy maps. To prove the theorem we have to show that the maximal ideals in the support of

$$K \ / \ \beta(H^1(\Gamma_0(Np^{r-1})))$$

are Eisenstein.

At this point we will use modular symbols. A convenient reference is [A-S]. We exposit briefly the theory of modular symbols in a form which is most useful to us. For this we consider $\mathbb{P}^1(\mathbb{Q})$ with the natural action of $SL_2(\mathbb{Q})$. We define D to be the free abelian group generated by $\mathbb{P}^1(\mathbb{Q})$, which again has a natural action of $SL_2(\mathbb{Q})$. We define D_0 to be the degree 0 subgroup of D, i.e., D_0 is defined by the following exact sequence:

$$0 \to D_0 \to D \to \mathbb{Z} \to 0,$$

where the map from D to $\mathbb{Z}$ is given by $\sum n_i P_i$ being mapped to $\sum n_i$. We consider the dual sequence:

$$0 \to \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, M) \to \mathrm{Hom}_{\mathbb{Z}}(D, M) \to \mathrm{Hom}_{\mathbb{Z}}(D_0, M) \to 0.$$

Here M is any module. For a congruence subgroup Γ of $SL_2(\mathbb{Z})$ (we consider M as a Γ module with trivial action), on taking the fixed points under the action of Γ of this sequence, the relevant part of the resulting long exact sequence is:

$$0 \to M \to \mathrm{Hom}_{\mathbb{Z}[\Gamma]}(D, M) \to \mathrm{Hom}_{\mathbb{Z}[\Gamma]}(D_0, M) \xrightarrow{\gamma} H^1(\Gamma, M) \to H^1(\Gamma, \mathrm{Hom}_{\mathbb{Z}}(D, M)). \tag{2}$$

We note that by Shapiro's lemma the last term of (2) is isomorphic to $\bigoplus_c H^1(U_c, M)$, where c runs over the cusps of the compactified curve associated to Γ, and U_c is the stabiliser of c in Γ. The map γ may be made explicit easily. One can check

that fixing $z \in \mathbb{P}^1(\mathbb{Q})$, $\gamma(f)(\tau) = f(\tau(z) - z)$. This does not depend on the choice of z. We also note that the above sequence is an exact sequence of Hecke modules. For the definition of the Hecke action on the terms of (2) we can again refer to [A-S]. (The reader may also look into [A-S, Proposition 4.2] where it is shown that $\mathrm{Hom}_\Gamma(D_0, M)$ is isomorphic to ${H_c}^1(Y_\Gamma, M)$, Y_Γ being the open curve associated to Γ and Γ a congruence subgroup of $SL_2(\mathbb{Z})$.) We see that the maximal ideals of the Hecke algebra in the support of the kernel and cokernel of γ are Eisenstein. Thus we conclude that:

$$H^1(\Gamma, M)_m \equiv \mathrm{Hom}_{\mathbb{Z}[\Gamma]}(D_0, M)_m \tag{3}$$

as Hecke modules, m being any maximal non-Eisenstein ideal of the Hecke algebra and by the subscript m we mean that we have localised at m. We denote the left-hand side of (3) by H_m. We call the right-hand side the space of modular symbols.

Using (3) we can now give a proof of Theorem 1. Suppose that (f, g) is in the kernel of α. So we have that:

$$f(x) + g(\pi x \pi^{-1}) = 0 \tag{4}$$

for all $x \in \Gamma_0(Np^{r+1})$. We set up some notation to ease the exposition. So define:

$$\Gamma := \Gamma_0(Np^r)$$

$$\Gamma' := \pi\Gamma_0(Np^r)\pi^{-1}$$

$$\Gamma'' := \Gamma \cap \Gamma'.$$

We define h_1 and h_2, elements of $H^1(\Gamma, M)$ and $H^1(\Gamma', M)$ respectively, by:

$$h_1(x) = -g(x), \ x \in \Gamma$$

and

$$h_2(\pi x \pi^{-1}) = f(x), \quad x \in \Gamma.$$

Using (4), we see that h_1 and h_2 restrict to $H^1(\Gamma'', M)$ to give the same element, which is the restriction of $-g$, as $h_1 - h_2 \in H^1(\Gamma'', M)$ restricts to $H^1(\Gamma_0(Np^{r+1})^0, M)$

to 0. (Here the superscript 0 denotes the subgroup consisting of elements whose upper right hand entry is divisible by p.) Then on using the injectivity of the restriction map in our present situation, we conclude that h_1 and h_2 restrict to the same element, namely $-g$, in $H^1(\Gamma'', M)$. Under these circumstances we have to examine the obstruction to the restriction of g to $H^1(\Gamma'', M)$ coming from the restriction of an element of $H^1(\Gamma_0(Np^{r-1}), M)$. We can use (3) to see that the obstruction is Eisenstein. The image of g in the H_m of (3) where m is any non Eisenstein maximal ideal, on using (3), with the group Γ'' being used for the Γ of (3), gives rise to a modular symbol which is invariant under the action of both Γ and Γ'. But $\Gamma_0(Np^{r-1})$ is generated by Γ and Γ' in $SL_2(\mathbb{Z})$. Thus, on using (3) again, we see that the image of g in H_m does come from restriction of an element of $H^1(\Gamma_0(Np^{r-1}), M)_m$. Thus we may consider the image in $H^1(\Gamma_0(Np^{r+1}), M)_m$ of ${\alpha_1}^* f + {\alpha_p}^* g$ as an element of $H^1(\Gamma_0(Np^r), M)_m$, and from (4), it restricts to 0 in $H^1(\Gamma_0(Np^{r+1}), M)_m$. Upon once again using the injectivity of restriction maps between H^1's of groups of varying p power level, we see that we have proven the theorem as it has been shown that the only maximal ideals in the support of A/J are Eisenstein.

Remark 2.

2.1 It is not true that $\Gamma_0(Np^{r-1})$ is isomorphic to the amalgamated product $\Gamma *_{\Gamma''} \Gamma'$ though this does map surjectively to it. One may in fact see that this amalgam (and even any quotient of it by a finite subgroup) cannot act discretely and faithfully on the upper half plane in general, as its H^2 which we may calculate by using the Lyndon exact sequence (see example on page 127 of [Se 2] for this), is too large for it to do so. We may also note that this H^2 is highly non-trivial as a Hecke module, containing as it does essentially all the information about p-new forms for the group $\Gamma_0(Np^{r+1})$. Thus Theorem 1 (or rather its proof) shows that these rather different groups have essentially the same abelianised quotients.

2.2 It is likely that Theorem 1 can be given a different proof using the method in Lemma 2.5 of [W] though the (p, p) case is explicitly excluded there.

2.3 We note that Theorem 1, in a sharper form, has been proven before in the special case when $N = 1$ and $r = 1$ in [L]. There it is proven that the kernel of the map in this special case is isomorphic to the Shimura subgroup. The method in [L] is completely different, uses reduction mod p of modular curves, and relies on the fact that $J_0(p)$ has toric reduction at p and hence perhaps will not work in the more general setting of Theorem 1.

2.4 One can easily see that in the setting of Theorem 1 too the anti-diagonal embedding of the Shimura subgroup lies in the kernel of α. We recall that the Shimura subgroup is by definition the kernel of the map $J_0(Np^r) \to J_1(Np^r)$ and in terms of group cohomology is associated to the image in $H^1(\Gamma_0(Np^r), M)$ (for suitable coefficients M) of homomorphisms of $(\mathbb{Z}/Np^r)^*$ under the inflation map from the quotient $\Gamma_0(Np^r)/\Gamma_1(Np^r)$ (see [L-S] for a more exact statement). From this description the claim follows as we note that if a homomorphism ξ of $\Gamma_0(Np^r)$ factors through $\Gamma_1(Np^r)$ then $(\alpha_p{}^*\xi)(g) = (\alpha_1{}^*\xi)(g)$ for any $g \in \Gamma_0(Np^{r+1})$. As Ling San pointed out to us, it is easy to see from Corollary 2 to Theorem 1 of [L-S], that the image of the anti-diagonal copy of the Shimura group is either trivial or of order p in A/J according as r is even or odd (assuming $r > 1$).

2.5 One can check that in the setting of Theorem 1, for $r > 1$, the finite group A/J is p-torsion. This follows from the shape of the matrix R of §1.3. From this one in fact checks that $\ker(R)/J$, which is easily seen to be a finite group, is p-torsion. We only briefly indicate the argument as we do not really need it in this chapter. R of §1.3 is nothing other than the composition $\alpha^\vee.\alpha$ where $\alpha^\vee$ is the dual map to α. From this we see that if (x, y) is in the kernel of R then $T_p x + py = 0$ (see proof of Theorem 2 in §1.3). We consider the image of (x, y) in $\ker(R)/J$. We need to show that this is annihilated by a power of p. For this we consider the image of x in $J_0(Np^r)/\alpha_p{}^*(J_0(Np^{r-1}))$ and write it as $\alpha_1{}^*(z) + w$ for $z \in J_0(Np^{r-1})$ and w in the p-new subvariety of $J_0(Np^r)$ (we are of course abusing notation and denoting an element of $J_0(Np^r)$ and its image in the quotient being considered by the same symbol). Then we see that the image of $T_p(x)$ in $J_0(Np^r)/\alpha_p{}^*(J_0(Np^{r-1}))$ is in $\alpha_1{}^*(J_0(Np^{r-1}))$. This follows from the fact that we are assuming $r > 1$ and

thus T_p annihilates the p-new subvariety of $J_0(Np^r)$. From this we deduce that A/J cannot have point of order ℓ for any prime ℓ different from p. It remains to determine exactly the p-torsion of the group of connected components of the kernel of α (which we have seen in Theorem 1 to be Eisenstein).

Remark 3. The method of proof of Theorem 1 yields the following result:

If α is the natural degeneracy map:

$$J_0(Np^r)^{n+1} \to J_0(Np^{r+n})$$

then the group of connected components of its kernel is Eisenstein. α is the sum of the maps ${\alpha_{p^i}}^*$ for $0 \leq i \leq n$ where ${\alpha_{p^i}}^*$ is the map induced by the map $\alpha_{p^i} : X_0(Np^{r+n}) \to X_0(Np^r)$ which is given by $\alpha_{p^i}(E, C_{Np^{r+n}}) = (E/C_{p^i}, C_{Np^{r+i}}/C_{p^i})$. Here as usual $C_{Np^{r+n}}$ is a cyclic subgroup of E of order Np^{r+n} and C_{p^i} is its subgroup of order p^i etc. We shall have use for this remark in §1.6.

In fact we state a more precise result after setting up some notation. For $1 \leq i < j \leq n+1$ we consider the abelian subvariety $A_{i,j}$ of $J_0(Np^r)^{n+1}$ given by the image of $J_0(Np^{r-j+i})$ under the map $x \to (0, \cdots, -{\alpha_{p^{j-i}}}^*(x), \cdots, {\alpha_1}^*(x), \cdots, 0)$ where the denoted non-zero entries are in the ith and jth place respectively. We use the convention that $A_{i,j}$ is 0 if $j - i > r$. We are of course continuing with our by now institutionalised abuse of notation by not distinguishing between what are morally the same degeneracy maps but which technically are different as they arise from different p power levels. We feel confident that this shall not cause any confusion. With this said, we can now state the generalisation of Theorem 1.

Theorem 1′. If A_n is the kernel of the above degeneracy map $\alpha : J_0(Np^r)^{n+1} \to J_0(Np^{r+n})$, then the sum of the $A_{i,j}$'s as above is the connected component of A_n and the group of connected components of A_n is Eisenstein.

Proof. The proof is quite similar to the proof of Theorem 1 and so we shall be brief. In fact the case $n = 1$, which is the case settled in Theorem 1, is the key case from which the others follow by a simple argument as we now indicate. We denote the sum of the $A_{i,j}$'s by J_n (which is easily checked to be in A_n) and claim as in

the statement of the theorem that J_n is the connected component of the identity of A_n. As in the proof of Theorem 1, for this it is enough to look at the degeneracy maps induced on the space of regular differentials. Then again by using the result in [A-L] we are done by using induction on n (the $n = 1$ case has been handled in the proof of Theorem 1). As we see that if $\sum_0^n f_i(p^i z) = 0$, then by [A-L], f_n is of level Np^{r-1}. Then by subtracting $(0, \cdots, -f_n(pz), f_n(z))$ from $(f_1, \cdots, f_n)$ we see by induction that our claim is true. Now we come to the statement about the group of connected components being Eisenstein. Thus if $(x_0, \cdots, x_n)$ is in the kernel of α we rewrite $\sum_0^n \alpha_{p^i}{}^*(x_i)$ as $\sum_0^{n-1} \alpha_{p^i}{}^*(x_i) + \alpha_{p^n}{}^*(x_n)$. Then we note that $\sum_0^{n-1} \alpha_{p^i}{}^*(x_i)$ lies in $\alpha_1{}^*(J_0(Np^{r+n-1}))$. On using Theorem 1 in the case of the degeneracy map $J_0(Np^{r+n-1})^2 \to J_0(Np^{r+n})$, we see that if $(x_0, \cdots, x_n)$ is in the non-Eisenstein part of the kernel, then $\alpha_{p^{n-1}}{}^*(x_n)$ which is in $J_0(Np^{r+n-1})$ arises by pull back by α_1 from $J_0(Np^{r+n-2})$. We see by iterating this process that x_n in $J_0(Np^r)$ is in fact in $\alpha_1{}^*(J_0(Np^{r-1}))$, i.e., $x_n = \alpha_1{}^*(y)$ for $y \in J_0(Np^{r-1})$. Now by subtracting $(0, \cdots, -\alpha_p{}^*(y), \alpha_1{}^* y)$ from $(x_0, \cdots, x_n)$ we see that we are done by induction on n (the case $n = 1$ being Theorem 1).

Remark 4. We also note an application of modular symbols in the raising of levels situation considered by Ribet in [R]. In this paper the crucial point was to calculate the kernel of the degeneracy map

$$\alpha : J_0(N)^2 \to J_0(Np).$$

By considerations similar to the preceding we see that the non-Eisenstein part of the kernel of α arises from an element of $\mathrm{Hom}_\Gamma(D_0, M)$ where Γ here is the analog of $\Gamma_0(N)$ in the group $SL_2(\mathbb{Z}[1/p])$. We would like to show that any Hecke eigenform in this space of modular symbols is forced to be Eisenstein. This follows from the congruence subgroup property which is known for $SL_2(\mathbb{Z}[1/p])$ (see [Se 3] for this). To see this implication we need only note that on using the exact sequence (2) made with Γ the subgroup of $\Gamma_0(N)$ type of $SL_2(\mathbb{Z}[1/p])$, we see that $\mathrm{Hom}_\Gamma(D_0, M)$ is Eisenstein as we know this to be true for $H^1(\Gamma, M)$ (for M as before considered as a trivial Γ module) as a direct consequence of the congruence subgroup property

enjoyed by $SL_2(\mathbb{Z}[1/p])$. Thus we can avoid using the description of Ihara (used in [R] and for which see [Se 2]) of $SL_2(\mathbb{Z}[1/p])$ as an amalgamated product.

1.3. Raising the level

We now come to the result about raising levels in the (p,p) case. We shall supplement this result in §1.5 by studying congruences between forms of level N, Np and Np^2 (mod p) which is ignored here. We restate and prove the theorem of the introduction.

Theorem 2. *Suppose f is a newform of weight 2 for the group $\Gamma_0(Np^r)$, $(N,p)=1$ and that the mod p representation f gives rise to is absolutely irreducible. Then for any $s > max(r,2)$, there exists a newform g, of weight 2 for the group $\Gamma_0(N'p^s)$, where $N'|N$, such that:*

$$f \equiv g \pmod{\mathfrak{p}},$$

where $\mathfrak{p}$ is a place above p.

Remark 5. We note that for large enough s our form g cannot arise from twisting from the form f as we are requiring that g be a form for the group $\Gamma_0(Np^s)$. (This fact is used crucially in the application of Theorem 2 to prove Theorem 5 in §1.7.) For if g were to arise by twisting, then the twist would have to be by a character whose conductor has the same radical as Np and then the resulting form cannot be on $\Gamma_0(Np^s)$ if the order of the character is greater than 2, as we may see by considering nebentypes. As there are only finitely many characters of this type, we see that for large enough s the form g the theorem produces cannot be a twist of f.

Remark 6. It is well known that there are only finitely many Hecke eigensystems mod p which can occur in cusp forms of weight 2 and level Np^∞ (by this we mean of level Np^r for varying r). In fact one knows that any such eigensystem (mod p) already occurs in $S_2(\Gamma_1(Np^2))$. For this we refer to the discussion in §2 and §3 of [R 2]. Thus one knows that there are many congruences mod p between newforms of weight 2 and varying p power level. Theorem 2 provides a more precise version of this qualitative fact. For instance after Theorem 2 one knows that f mod p occurs

in the p-new quotient of $S_2(\Gamma_0(Np^s))$ for infinitely many s. The general fact we have quoted does not seem to give this (but see Remark 8 below).

In Theorem 2 by a congruence between modular forms, we mean a congruence of the Fourier coefficients outside Np and in general throughout this chapter by a congruence mod p between newforms we shall mean that they give rise to isomorphic (absolutely irreducible) mod p representations.

Proof of Thorem 2. We follow the method initiated by Ribet in [R]. There the general strategy to prove that a maximal ideal of the Hecke algebra $\mathbb{T}_{Np^s}$ is a prime of fusion, i.e., arises by pull back from the p-old and p-new quotients of $\mathbb{T}_{Np^s}$, is to produce a Hecke module for which the Hecke action factors through the p-old and p-new quotients and then show that the maximal ideal in question is in the support of this module. A natural candidate for such a module is the (finite) module given by the intersection of the p-old and p-new subvarieties of $J_0(Np^s)$. Thus we need to be able to calculate, in the light of this strategy, the maximal ideals in the support of this module.

For this we observe that because of the autoduality of the Jacobian, the natural degeneracy map

$$\alpha : J_0(Np^{s-1})^2 \to J_0(Np^s)$$

induces a map :

$$\alpha^\vee : J_0(Np^s) \to J_0(Np^{s-1})^2.$$

By Theorem 1 we know that the group of connected components of the kernel of α is Eisenstein. This entails that the group of connected components of the kernel of $\alpha^\vee$ is Eisenstein, as these two groups of connected components are in Cartier duality with each other. This general fact about abelian varieties can be seen by an argument which we owe to Ribet. As we note that if $T : A \to B$ is a map of abelian varieties, then it can be viewed as taking place in 3 stages. At the first stage one considers the quotient of A by the connected component of the identity of the kernel of T. At the second stage one considers the quotient by the image in this of the (finite) group of the connected components of the kernel of T. At the third stage one considers the inclusion of A / $\ker(T)$ into B. Then the kernel of the dual

map $T^\vee : B^\vee \to A^\vee$ may be unscrewed corresponding to the above 3 stages. The kernel of the dual map corresponding to the third stage is the connected component of the identity of the kernel of $T^\vee$. The kernel of the map dual to the second stage is the group of connected components of the kernel of $T^\vee$. The kernel of the dual map corresponding to the third stage is 0. From this the result follows as it is well known that the kernel of an isogeny is in Cartier duality with the kernel of its dual map. From this it follows that the group of connected components of the kernel of $\alpha^\vee$ is Eisenstein as the Hecke operators T_n for $(n, Np) = 1$ are self adjoint under this pairing.

The connected component of the identity of the kernel of $\alpha^\vee$ is by definition the p-new subvariety of $J_0(Np^s)$. By a well known computation we see that the composition $\alpha^\vee.\alpha$ when written as a 2 by 2 matrix of endomorphisms of $J_0(Np^{s-1})^2$ is given by:

$$R = \begin{pmatrix} p & T_p^* \\ T_p & p \end{pmatrix}.$$

$T_p{}^*$ is the image of T_p under the Rosati involution of the Jacobian. To be explicit, T_p for us is defined by $(\alpha_p)_*.\alpha_1{}^*$, which induces the usual operator T_p on cusp forms (whose level is divisible by p). (We believe there is a small error in §13 of [M-R] (which of course does not affect anything in their paper) where the operator given by $(\alpha_1)_*.\alpha_p^*$ is called T_p and it is stated that $T_p.\alpha_1{}^* = \alpha_1{}^*.T_p$ with T_p on either side of the equation denoting the action at the appropriate level. But it may be checked that, denoting this operator by the neutral symbol B, $B.\alpha_1^* = p.\alpha_p^*$ where α_1^* and α_p^* are maps from $J_0(Np^{s-2})$ to $J_0(Np^{s-1})$. This contradicts the property claimed for what is called T_p in §13 of [M-R]. This is one place where our conflation of degeneracy maps is likely to cause vertigo! We labour this point as it has confused us). The description of R results from this description of the Hecke operators using degeneracy maps, and on noting that the maps α_1 and α_p have degree p. We remark that with our conventions, $T_p.\alpha_1{}^* = \alpha_1{}^*.T_p$ with T_p referring to the appropriate action at the respective levels (we remind ourselves that by the assumptions in Theorem 2, $s - 1 > 1$). We note that, in the case when $r \geq 2$, a

newform f of weight 2 for the group $\Gamma_0(Np^r)$, where $r \geq 2$, is annihilated by T_p for the well known reason that with our assumption that $r \geq 2$, T_p decreases the level of f. For the case $r \leq 1$ of the theorem, we produce from f a p-old form f'' of level Np^2 whose L series has no Euler factor at p, but whose Euler factors outside p agree with those of f. Thus the maximal ideal we are interested in proving to be a prime of fusion, is the maximal ideal of residual characteristic p associated to either f or f'' of ${\mathbb{T}_{Np^s}}^{(Np)}$ (where by this we mean the Hecke algebra deprived of the Hecke operators T_n where (n, Np) is not 1). Having said this we shall now only consider the case $r \geq 2$ of the theorem as the other case is entirely similar. We shall denote this maximal ideal corresponding to f by m. On taking this into account and noting the shape of the matrix R we can write down p torsion elements of the abelian variety ${A_f}^2$ associated to f of the form $(x, 0)$ (see [S, Theorem 7.14] for this association; for purposes of orientation we may recall that A_f is a subvariety of $J_0(Np^{s-1})$ whose space of regular differentials is spanned by f and its Galois conjugates under the standard Galois action on q-expansions), which lie in the kernel of R, but are not in the kernel of α, and such that their images under α are not in the group of connected components of the kernel of $\alpha^\vee$.

Namely, we see that if V is a two-dimensional vector space over ${\mathbb{T}_{Np^{s-1}}}^{(Np)}/m$ (we are identifying m and its image in ${\mathbb{T}_{Np^{s-1}}}^{(Np)}$) which affords the representation ρ_m (see §1.2) and occurs in $J_0(Np^{s-1})$, then $(V, 0)$ is mapped injectively to the intersection of the p-old and p-new subvarieties of $J_0(Np^s)$. It is mapped injectively by inspection as we have noted in §1.2 that the map ${\alpha_1}^*$ is injective. It is mapped to the intersection as firstly by what we have noted about the behaviour of T_p it does lie in the kernel of R. Secondly the image of $(V, 0)$ under α has trivial intersection with the group of connected components of $\alpha^\vee$ as V is absolutely irreducible as a Galois module and we have shown earlier that this group of connected components is Eisenstein. (It may be useful to note that the kernel of R has the following filtration:

$$0 \subseteq \Omega_1 \subseteq \Omega_2 \subseteq \Omega_3 \subseteq \ker(R).$$

The successive quotients here are identified respectively with $\beta(J_0(Np^{s-2}))$, the group of connected components of α, the intersection of the p-old and p-new subva-

rieties of $J_0(Np^s)$ and the dual of the group of connected components of α.) This then shows that m is a prime of fusion as it is in the support of the intersection of the p-old and p-new subvarieties which follows from what we have just said that this intersection contains a mod p Galois representation isomorphic to ρ_m. Thus we obtain the result claimed in the theorem.

Remark 7. We note that in the context of Theorem 2 the only prime with respect to which one can raise the p power level of the newform f is p. (In this remark by congruence between cusp forms we mean term by term congruence of all the Fourier coefficients. This makes good sense in the situation of $r > 1$ of Theorem 2, as the T_n of higher level restricts to give T_n of lower level and T_p annihilates newforms in $S_2(\Gamma_0(Np^s))$ for $s > 1$ as we remarked earlier. It may be checked that in the setting of Theorem 2 for the case $r > 1$ we can raise the levels in this stronger sense to get a form g for the group $\Gamma_0(Np^s)$ which is an eigenform for the Hecke operators at level Np^s, and whose minimal p power level is p^s.) This may be seen on using the characterisation of newforms of Serre as the kernel of the intersections of the natural trace maps, noting that the trace map preserves the integrality of the Fourier coefficients outside p and the fact that the degree of the covering $X_0(Np^{r+1}) \to X_0(Np^r)$ is p. (We also refer the reader to Remark 2.5 of §1.2.)

Remark 8. We discuss the relation between Theorem 2 and twisting. As per Remark 5, for large enough s, the form g the theorem provides cannot arise by twisting. We now assume $p \geq 5$. We may apply Carayol's lemma, cf. [C], to produce from a newform f in $S_2(\Gamma_0(Np^r))$ a Hecke eigenform h in $S_2(\Gamma_0(Np^s), \varepsilon)$ congruent to f mod a place above p, for $s \geq r$, where ε is any character of order a power of p and conductor dividing p^s which can be specified in advance. If we consider the extreme case when the conductor of ε is p^s ($s \geq 2$), then we know that the automorphic representation corresponding to h is principal series at p. In this case we may twist h by ε' where ε' has p power order and is such that $(\varepsilon')^2 = \varepsilon^{-1}$ (we are assuming that $p > 2$, so this is possible). In this case the resulting form, say $h_{\varepsilon'}$, is easily checked to be in the p-new part of $S_2(\Gamma_0(Np^{2s}))$. Thus we see

that the use of Carayol's lemma gives a weaker result than Theorem 2 (but which is still sufficient for the application to Theorem 5) which will say that for any $s \geq r$ (and $s \geq 2$), f occurs mod p in the p-new part of $S_2(\Gamma_0(Np^{2s}))$. In fact this lemma of Carayol does help us in further understanding, in conjunction with Theorem 2, congruences in the (p,p) case as we point out in Theorem $4'$ of §1.5.

1.4. Lowering the level

In this section we discuss the lowering of levels in the (p,p) situation. Namely, given a newform f of weight 2 on $\Gamma_0(Np^r)$ we would like to find the minimal p power level at which it occurs, i.e., we would like to find the least s such that $S_2(\Gamma_0(Np^s))$ contains a form which is congruent to f mod p. We will use the terminology that f occurs in $S_2(\Gamma_0(Np^s))$ to signify this and we shall call the least such s the minimal p power level of f (so s is in fact the minimal exponent of p in any level in which f occurs mod p in the space of weight 2 cusp forms). Throughout this section we assume that the mod p representation attached to f is absolutely irreducible and that $p \geq 5$. We state a proposition which settles this question in terms of the behaviour of the mod p representation attached to f when considered as a representation of the decomposition group at p.

Proposition 1. *The minimal p power level of f is either 0, 1 or 2. We can determine this minimal level in terms of the following trichotomy:*

(a) If the mod p representation attached to f (which we denote by $\rho_{f,p}$) is finite at p, then the minimal p power level of f is 0.

(b) If $\rho_{f,p}$ is not finite at p but as a representation for D_p (the decomposition group at p) has a one-dimensional unramified quotient, then the minimal p power level of f is 1.

(c) If neither is the case then the minimal p power level of f is 2.

(For the notion of finiteness of ρ at p we refer to [Se].)

Proof. The proof just consists of quoting results from [E] and [Se]. That the

minimal p power level of f is at most 2 follows from the fact, see [R 2] and Remark 6 of §1.3 of the present chapter, that f occurs in $S_2(\Gamma_0(N) \cap \Gamma_1(p^2))$. Then we may in fact see that f occurs in $S_2(\Gamma_0(Np^2))$ on using a well known lemma of Carayol about lifting of nebentypes, cf. [C].

Now we use the results in [E] and [Se] to determine exactly the minimal level. We first note that it is well known that f occurs in $S_k(\Gamma_0(N))$ for some k. For the sake of being more self contained we recall briefly the conjecture of Serre about the least weight from which an absloutely irreducible modular representation arises, from a level N which is prime to p. This conjecture is proven by Edixhoven in [E]. In our proof of the proposition the only case of this conjecture of Serre which is relevant is the case when the mod p representation we are considering is reducible when restricted to the decomposition group D_p at p. Thus we consider an absolutely irreducible mod p representation associated to a newform g in $S_k(\Gamma_1(N))$ for N, as usual, prime to p. We denote this by $\rho_{g,p}$. Then Serre attaches to this an integer $k(\rho)$ (≥ 2) which is now proven in [E] to be the least weight k' (≥ 2) at which $\rho_{g,p}$ arises from a form in $S_{k'}(\Gamma_1(N))$. Serre's invariant $k(\rho)$ only depends upon the restriction of $\rho_{g,p}$ to D_p, the decomposition group at p (in fact it depends only on the restriction to the inertia group at p). Now we assume that this restriction is not irreducible as that is the only case we need to look at here. Then $\rho_{g,p}|D_p$ is of the form:

$$\begin{pmatrix} \chi^{\beta}\varepsilon_1 & * \\ 0 & \chi^{\alpha}\varepsilon_2 \end{pmatrix}$$

where χ is the mod p cyclotomic character and ε_i are unramified characters. α and β are well defined mod $p-1$. We assume first that the wild part of the inertia group at p acts trivially and in this case normalise so that $0 \leq \alpha \leq \beta \leq p-2$. Then $k(\rho)$ is defined to be $1+p\alpha+\beta$ if $(\alpha,\beta) \neq (0,0)$ and p otherwise. In the other case when the wild inertia group does not act trivially one normalises so that $0 \leq \alpha \leq p-2$ and $1 \leq \beta \leq p-1$ and then defines a and b to be the maximum and minimum of α and β respectively. Then if β is not $\alpha+1$, Serre sets $k(\rho)$ in this case equal to $1+pa+b$. In the case when $\beta = \alpha+1$, $k(\rho)$ is defined to be $1+pa+b$ if ρ is peu ramifié at p and $1+pa+b+p-1$ otherwise. For the notion of peu ramifié, as for all of this paragraph, we refer to [Se]. Now we go back to the proof of the proposition.

We see that by Proposition 4 of [Se] and the main theorem of [E] a necessary and sufficient condition for the minimal level of f to be 0 is that ρ (as we shall call $\rho_{f,p}$) is finite at p (this step in [E] uses crucially Mazur's principle, for which see [R 1]). This follows from these results in [Se] and [E] on noting that the determinant character of ρ is the mod p cyclotomic character χ as then it is predicted in [Se] as we have recalled, and proven in [E], under our further hypothesis that ρ is finite at p, that $k(\rho)$ is 2. This proves part (a) (in an if and only if form) of the proposition.

So we may now assume that ρ is not finite at p. Then a necessary condition for f to have minimal p power level equal to 1 is that ρ when restricted to D_p should have the form:

$$\rho|D_p = \begin{pmatrix} \chi\varepsilon_1 & * \\ 0 & \varepsilon_2 \end{pmatrix} \qquad (*)$$

where ε_i for $i = 1, 2$ are unramified characters. This follows from Deligne's theorem (cf. Theorem 2.5 of [E]) on noting that if f has minimal p power level 1 then it is ordinary at p and that the weight filtration of an eigenform in $S_2(\Gamma_0(Np))$ is $\leq p+1$. (In fact it is either 2 or $p+1$ as a consequence of [A-S]. By the weight filtration of an eigenform we shall mean the least weight at which it arises mod p from a level prime to p, i.e., after [E], the $k(\rho)$ of the corresponding mod p representation ρ). Thus a necessary condition for f to have minimal level 1 is that the mod p representation should have a one-dimensional unramified quotient. Conversely if $(*)$ holds, it was conjectured in [Se] (see pg. 187) as we have recalled, and is proven in [E], that $k(\rho)$ is then either 2 or $p+1$ and it is 2 if and only if ρ is either completely reducible when restricted to the inertia group at p or peu ramifié at p. This latter condition, in the case when $\rho|D_p$ is of the form $(*)$, is shown to be equivalent to ρ being finite at p in §8 of [E]. Thus under our assumption that ρ is not finite at p it follows that f occurs in $S_{p+1}(\Gamma_0(N))$ but not in $S_2(\Gamma_0(N))$. But now from a result of Ash and Stevens [A-S, Theorem 3.5] we see that f occurs in $S_2(\Gamma_0(Np))$. Thus we have proven part (b) of the proposition. On taking into account the fact that we started the proof by quoting, i.e., that the minimal p power level of f is ≤ 2, we see that we have proven the proposition.

Remark 9. We note that using the results of [E] and [A-S] one may settle the

question of the minimal p power level of a form f more generally, i.e., assuming that f occurs in $S_2(\Gamma_1(Np^r))$ rather than the more restrictive assumption that it occurs in $S_2(\Gamma_0(Np^r))$. We state the result. We will of course by the minimal p power level mean the least s such that, mod p, f occurs in $S_2(\Gamma_1(Np^s))$. Denoting the associated mod p representation by ρ we can determine the minimal level in terms of the invariant $k(\rho)$ of Serre. (Here we note that the $k(\rho)$ in [Se] is never 1 while in [E] it can be 1. We use the original definition of Serre.) We deduce from [E] and [A-S] that the minimal level of f is 0 if $k(\rho)$ is 2, is 1 if $2 < k(\rho) \leq p+1$ and is 2 if $k(\rho) > p+1$. By [C] (we recall that we are assuming $p \geq 5$) we also note that the nebentype with which f occurs in the minimal level may be taken to have order prime to p. Proposition 1 is a particular though more self-contained case of this.

1.5. Congruences between forms of level N, Np and Np^2

The study of congruences between eigenforms of level N, Np and Np^2, which has been avoided in Theorem 2, can also be carried out. Congruences between forms of level N and level Np (even mod p) can be handled by the methods in [R]. One may analyse congruences between forms of level N and Np^2 and between those of level Np and Np^2 using our method. The case of congruences between forms of level Np and Np^2 is somewhat delicate. Throughout we are continuing with our assumption that $p \geq 5$ and that the mod p representation attached to the forms we shall consider is absolutely irreducible. As usual by a congruence mod p we shall mean that the associated Galois representations of residue characteristic p are isomorphic. We now state and prove a theorem which studies congruences between forms of low p power levels.

Theorem 3. *We can classify congruences between forms of level N, Np and Np^2 as follows:*

1. If f is a newform in $S_2(\Gamma_0(N))$ then a necessary and sufficient condition for there to be a congruence (mod p) between f and a p-new form in $S_2(\Gamma_0(Np))$ is that $a_p(f) \equiv \pm 1 \pmod{\mathfrak{p}}$ where $\mathfrak{p}$ is a place above p as usual.

2. If f is a newform in $S_2(\Gamma_0(N))$ then f is always congruent (mod p) to some form in $S_2(\Gamma_0(Np^2))$ which is p-new.

3. If f is a newform in $S_2(\Gamma_0(Np))$ then f is congruent (mod p) to a p-new form in $S_2(\Gamma_0(Np^2))$ if and only if $\rho_{f,p}$ is finite at p.

Proof.

Case 1. The sufficiency follows from [R] as is remarked upon in [R 3]. The necessity is seen by using Deligne's theorem that the mod p representation attached to a form of weight 2 which has level divisible exactly by p, when restricted to the decomposition group at p, has an unramified quotient on which it acts by a character of order dividing 2 (see §1.4). Thus if the level of f can be raised as above, the corresponding mod p representation is reducible at p and hence on using Theorem 2.5 and 2.6 of [E] (see also §1.4) we see that the above condition is necessary as the cited theorems in [E] force the eigenvalue $a_p(f)$ to be ± 1 modulo $\mathfrak{p}$. (We are grateful to K. Ribet for a helpful discussion about this case.)

Case 2. The level of a newform f of weight 2 and level N mod p can always be raised to get a congruent newform of weight 2 and level $N'p^2$ where $N'|N$. This follows from the procedure of the proof of Theorem 2 as from f we can produce a p-old form f' of level Np (the Euler factors of whose L series outside p agree with those of f) which is not ordinary at p. Then again the form of the matrix R in §1.3 in the case of going from level Np to Np^2 proves what we claimed.

Case 3. Now we have to deal with the case when f is a newform in $S_2(\Gamma_0(Np))$. If the corresponding mod p representation is finite at p (see [Se] and §1.4) then the level of f can be raised to get a newform g congruent to f mod p and level divisible by p^2 as by Proposition 1 of §1.4 we see that f is congruent to a form in $S_2(\Gamma_0(N))$ and then we use Case 2 for going from level N to Np^2.

If the mod p representation attached to f is not finite at p then we have to prove that the p level of f mod p cannot be increased to p^2. We note that Theorem 2 shows that one can always raise the level to get a newform g of level divisible by p^r for any $r > 2$.

So we now suppose that we have a newform g of weight 2 for $\Gamma_0(Np^2)$ such that the corresponding mod p representation is not finite at p. Then in order to prove Case 3 of the theorem we need to prove that the p part of the level of g cannot be lowered. As otherwise g would be congruent to a form in $S_2(\Gamma_0(Np))$ which is new at p (as $\rho_{g,p}$ is not finite at p). From this it follows that the mod p representation attached to g is reducible when restricted to the decomposition group at p (by a theorem of Deligne used above). We may also assume that the $\rho_{g,p}|D_p$ (D_p is the decomposition group at p) is not completely reducible. As otherwise it will have the form:

$$\begin{pmatrix} \chi^a\varepsilon_1 & 0 \\ 0 & \chi^b\varepsilon_2 \end{pmatrix}.$$

Now if neither of a and b is congruent to 0 mod $p-1$ we get that g (by [E]; see also §1.4) cannot occur mod p in any $S_k(\Gamma_0(N))$ for $k \leq p+1$ and hence cannot occur in $S_2(\Gamma_0(Np))$ by a result in [A-S] that we have used often already. If only one of a or b has non-zero residue class mod $p-1$, say a, then we see by looking at determinant character of $\rho_{g,p}$ that $a \equiv 1 \pmod{p-1}$. But then $\rho_{g,p}$ is finite at p (see Proposition 4 of [Se]). Thus from now we assume that $\rho_{g,p}|D_p$ is not completely reducible.

Now we consider two cases. One is when the p component of the automorphic representation corresponding to g is a principal series or twist of a special representation at p. Then we claim that g arises as a twist from a form h in $S_2(\Gamma_1(Np))$ by a primitive character of conductor p. For this it is enough to look at the case when the p component of π_g is principal at p. It is then of the form $\pi(\xi_1, \xi_2)$ where ξ_i are quasicharacters of $\mathbb{Q}_p$ of conductor p. This has to be the case by our assumption on the nebentype of g. This justifies the claim. (We note that by using the lemma of Carayol, cf. [C], any newform in $S_2(\Gamma_0(N) \cap \Gamma_1(p^2))$ already occurs mod p in $S_2(\Gamma_0(Np^2) \cap \Gamma_1(p))$ and then again we may again argue as we have done.)

Using [A-S] (Theorem 3.4 and 3.5) we know that h mod p has weight filtration $\leq p+1$. Then on using this and Theorems 2.5 and 2.6 of [E] together with our assumption that the mod p representation associated to g is reducible, we see that

the mod p representation associated to h when restricted to D_p is given by:

$$\begin{pmatrix} \chi^a \varepsilon_1 & * \\ 0 & \varepsilon_2 \end{pmatrix} \qquad (*)$$

where ε_i, for $i = 1, 2$, as before are unramified characters and χ is the mod p cyclotomic character and a has non-zero residue class mod $p-1$. This follows because from the definition of the invariant $k(\rho)$ of Serre, $\rho_{h,p}|D_p$ has either an unramified quotient or an unramified subspace (as the filtration of h mod p is $\leq p+1$). But from Deligne's theorem (Theorem 2.5 of [E]) it has an unramified quotient. We note that the determinant character of $\rho_{h,p}$ is ramified at p (as the determinant character of $\rho_{g,p}$ is χ and $p > 2$). Thus we see that as $\rho_{h,p}|D_p$ is not completely reducible, it cannot have an unramified subspace and hence is of the form $(*)$.

Now our assumption that g, mod p, occurs in $S_2(\Gamma_0(Np))$ fetches us a contradiction as we claim that then it has both an unramified quotient and an unramified subspace which contradicts the fact that $\rho_{g,p}|D_p$ is not completely reducible (we are using here the fact that the determinant character of $\rho_{g,p}$ is ramified at p). The claim is true because by Deligne's theorem quoted above the mod p representation attached to a p-new form in $S_2(\Gamma_0(Np))$ has an unramified quotient when restricted to D_p. But as g arises from h by twisting by a primitive character of conductor p and as the filtration of g is $\leq p+1$ (by our assumption that, mod p, g occurs in $\Gamma_0(Np)$; in fact under this assumption it has filtration exactly $p+1$ as the corresponding representation is not finite at p, cf. §1.4), we see from the definition of $k(\rho)$ in [Se] and $(*)$ that the twist has to be by χ^{-a}. From this it follows that $\rho_{g,p}$ when restricted to D_p has an unramified 1 dimensional subspace. This completes the treatment of the case when π_g is principal or twist of special at p.

(It is possible to give a different argument than the one above to prove what we want in the principal series case and which is similar to the argument we are about to give in the supercuspidal case below. As, in the case when $\pi_p(g)$ is principal, the abelian variety associated to g will acquire good reduction, by virtue of a theorem of Langlands and Carayol, over $\mathbb{Q}(\mu_p)$. This extension is tamely ramified at p and then

arguing as we do below in the supercuspidal case, we are done. This alternative argument in the principal series case was pointed out by D. Prasad after having seen an earlier version of our proof of Theorem 3.)

Now we deal with the other case, i.e., when the p component of π_g is supercuspidal at p. Thus assuming that π_g is supercuspidal at p and the p part of its conductor is p^2 we see that the p component of π_g is the Weil representation associated to a primitive character, say ξ, of conductor p of the unramified quadratic extension of $\mathbb{Q}_p$ which does not factor through the norm character (we recall that we are assuming that p is greater than 2 and refer for instance to the cuspidal case in [C] for more details on this). As we are in the (p,p) situation we cannot use the Langlands and Carayol theorem to deduce from this information about the p-adic representation associated to g when restricted to D_p. But we may use the theorem of Carayol, cf. [C 1], in conjunction with the Néron-Ogg-Shafarevich criterion to conclude that $\rho_{g,p}$ becomes finite at a place above p of K for K a tamely ramified extension of $\mathbb{Q}_p$ (as the abelian variety associated to such a g acquires good reduction over an extension K of this type). But the mod p representation associated to a form f in $S_2(\Gamma_0(Np))$ which is not finite at p (i.e., is not associated to a finite flat group scheme over $\mathbb{Z}_p$) remains so even when restricted to the ring of integers of any tamely ramified extension of $\mathbb{Q}_p$. As we see (see the discussion in Proposition 13.2 of [Gr]) that after making an unramified base change from $\mathbb{Z}_p$ to the ring of integers R of an unramified extension L of $\mathbb{Q}_p$, the Galois module V affording $\rho_{f,p}$ when restricted to $D_{\mathfrak{p}}$ (for $\mathfrak{p}$ a place above p) has the form:

$$0 \to \mu_p \otimes E^\vee \to V \to \mathbb{Z}/p \otimes E \to 0$$

where E is the finite extension of $\mathbb{F}_p$ which is given by $\mathbb{T}_{Np}/m$, m being the maximal ideal associated to f. This follows from the theorem of Deligne (Theorem 2.5 of [E]). By Kummer theory such an extension corresponds to an element in $L^*/L^{*p} \otimes E^\vee$. Here by definition $E^\vee$ is $\mathrm{Hom}(E, \mathbb{Z}/p)$. As V is not finite at p the class of the extension does not come from $R^*/R^{*p} \otimes E^\vee$ (as such extensions over R correspond to a class in $\mathrm{Ext}^1{}_R(E, E^\vee \otimes \mu_p)$ which by Kummer theory, as R is a discrete valuation ring, is isomorphic to $R^*/R^{*p} \otimes E^\vee$). But then it follows from this that the above

extension will not correspond to a finite, flat extension even after making a base change to the ring of integers of a tamely ramified extension of $\mathbb{Q}_p$. As we see that as V is not finite at p, at least one extension of the form:

$$0 \to \mu_p \otimes E^\vee \to V' \to \mathbb{Z}/p \to 0$$

with V' a Galois submodule of V gives rise to a class in $H^1(L, \mu_p \otimes E^\vee) = L^*/L^{*p} \otimes E^\vee$ which does not come from $H^1{}_{f.p.p.f.}(R, \mu_p \otimes E^\vee) = R^*/R^{*p} \otimes E^\vee$. The corresponding statement will remain true even after going to the ring of integers of a tamely ramified extension of $\mathbb{Q}_p$.

(We need to go through this additional step as the E action need not extend to the maximal finite flat extension associated to V over the ring of integers of K as the ramification index of K could well be $\geq p-1$; see [Ra]. We are using in the above the following general fact (for which again the reference is [Ra]):

Let $G = \mathrm{Gal}(\overline{K}/K)$ and let

$$0 \to X \to Y \to Z \to 0$$

be an exact sequence of G-modules. If Y is the generic fibre of a finite flat group scheme $\mathcal{Y}$ over the ring of integers of K then there are unique finite flat group schemes $\mathcal{X}$ and $\mathcal{Z}$ such that this sequence is the generic fibre of the following exact sequence of finite flat group schemes over the ring of integers of K:

$$0 \to \mathcal{X} \to \mathcal{Y} \to \mathcal{Z} \to 0.$$

As another reference for the above we may refer the reader to Appendix B of Milne's book [Mil]. We are grateful to D. Prasad for pointing out this reference.)

Thus we have proven that the p part of the level of a p-new form in $S_2(\Gamma_0(Np^2))$ which is supercuspidal at p cannot be lowered if the corresponding mod p representation is not finite at p. This completes the proof of Theorem 3.

Theorem 2, the result in [C] about lifting nebentypes (cf. Remark 8), the results of §1.4 and Theorem 3 yield a fairly complete analysis of congruences between cusp

forms in the (p,p) case. We can sum up these discussions in the form of the following theorem.

Theorem 4. *If f is a newform in $S_2(\Gamma_0(Np^s))$, then we may determine the different p power levels at which f (mod p) occurs as follows (we assume as usual that $\rho_{f,p}$, the mod p representation attached to f, is absolutely irreducible):*

1. If $\rho_{f,p}$ is finite at p then f occurs mod p in $S_2(\Gamma_0(Np^r))^{p-new}$ for all $r \geq 0$, except possibly $r = 1$. It occurs in $S_2(\Gamma_0(Np))^{p-new}$ if and only if $a_p(g) \equiv \pm 1$ (mod $\mathfrak{p}$) for $\mathfrak{p}$ a place above p and g in $S_2(\Gamma_0(N))$ a newform congruent to f mod p.

2. If $\rho_{f,p}$ is not finite at p, but if the mod p representation restricted to D_p has an unramified quotient then f occurs (mod p) in $S_2(\Gamma_0(Np^r))^{p-new}$ for all $r \geq 1$ except $r = 2$. It does not occur in $S_2(\Gamma_0(Np^2))^{p-new}$ and nor does it in $S_2(\Gamma_0(N))$.

3. If $\rho_{f,p}$ is neither finite at p nor does it have an unramified quotient when restricted to D_p, then f occurs mod p in $S_2(\Gamma_0(Np^r))^{p-new}$ if and only if $r \geq 2$.

Proof. This is just a summation of Theorem 2, Proposition 1 and Theorem 3.

Remark 10.

10.1 In some cases one can obtain finer information than is stated in Theorem 3. For instance if f is a newform in $S_2(\Gamma_0(N))$ and if its associated mod p representation is absolutely irreducible, but non semisimple when restricted to D_p, then one may check from the proof of Theorem 3 that the congruent form g which Theorem 3 will provide us which is p-new in $S_2(\Gamma_0(Np^2))$, is supercuspidal at p. This follows from the proof of Case 3 of Theorem 3 (more accurately the principal or twist of special subcase) on noting that if it were not supercuspidal the mod p representation aatched to g would be completely reducible when restricted to D_p (as it would have an unramified quotient and subspace).

There are many f's as above, i.e., such that the restriction of the mod p representation to D_p is non semisimple: for example see remark on page 274 of [Se 1] where it is remarked that it is expected (or more precisely the author of [Se 1] asks

a question which would have this as the affirmative answer) that most primes p at which an elliptic curve has good, ordinary reduction have the property that the mod p representation associated to the elliptic curve is non-semisimple when restricted to D_p.

This remark also gives a systematic way to construct irreducible mod p representations which come from forms which are supercuspidal at p and such that their restriction to D_p is not semisimple. H. Hida has pointed out to us that it should also be remarked that the mod p representation attached to a form $f \in S_2(\Gamma_0(N))$ can be irreducible when restricted to D_p, even though in this case $\pi_p(f)$ is principal. As by the theorem of Fontaine and Serre (Theorem 2.6 in [E]), if the mod p representation attached to a newform f in $S_2(\Gamma_0(N))$ is irreducible, then if $a_p(f)$ is 0 modulo a place above p, the restriction of the corresponding mod p representation to D_p is irreducible.

10.2 We may also remark that the oft cited lemma of Carayol, cf. [C], along with the results in [E] can easily settle the question of determining the various p power levels at which a form can appear when we switch to the Γ_1 type situation. But we may also try to refine the analysis and even analyse the various combinations of level and nebentypes with which the form will appear mod p (we are grateful to Fred Diamond for suggesting that this issue should also be resolved). We state the result. In the following by a character we shall always mean a primitive character.

Theorem 4'. *Let f be a newform of weight 2 for the group $\Gamma_0(N) \cap \Gamma_1(p^t)$. Then the various p-power levels and nebentypes with which f occurs mod p may be determined as follows (we as usual assume that the mod p representation associated to f is absolutely irreducible):*

1. $k(\rho_{f,p}) = 2$: If $s > 1$ then f occurs mod p in the p-new part of $S_2(\Gamma_0(Np^s), \psi)$, where ψ is any character of conductor p^r $(r \leq s)$ and of order a power of p.

f occurs in the p-new part of $S_2(\Gamma_0(N) \cap \Gamma_1(p))$ if and only if $a_p(g) \equiv \pm 1 \pmod{\mathfrak{p}}$ for g a newform in $S_2(\Gamma_0(N'))$, for $N'|N$, congruent to f mod p and $\mathfrak{p}$ a place above p.

2. $2 < k(\rho_{f,p}) \leq p+1$ *: If* $s > 2$ *then* f *occurs mod* p *in the* p*-new part of* $S_2(\Gamma_0(Np^s), \psi)$ *where* $s > 2$ *and* ψ *is any character of conductor* p^r $(r \leq s)$ *which mod* p *is congruent to the nebentypus character of* f*.*

For the case when $s \leq 2$*,* f *occurs mod* p *in the* p*-new part of* $S_2(\Gamma_0(Np^2), \psi)$ *for any* ψ *of conductor dividing* p^2 *if* $2 < k(\rho_{f,p}) < p+1$*. If* $k(\rho_{f,p}) = p+1$*, then* f *occurs mod* p *in the* p*-new part of* $S_2(\Gamma_0(Np^2), \psi)$ *if and only if the conductor of* ψ *is* p^2 *and* ψ *has order* p*.* f *occurs in* $S_2(\Gamma_1(Np))$*. It does not occur in* $S_2(\Gamma_0(N))$*.*

3. $k(\rho_{f,p}) > p+1$ *:* f *occurs mod* p *in the* p*-new part of* $S_2(\Gamma_0(Np^s), \psi)$ *where* $s \geq 2$*,* ψ *is any character of conductor* p^r $(r \leq s)$ *which mod* p *is congruent to the nebentypus character of* f*. It does not occur in* $S_2(\Gamma_0(N) \cap \Gamma_1(p))$*.*

Proof. To justify this we apply Remarks 8 and 9 and Theorem 2 of this paper in tandem with Theorem 3.1 of [A-Li].

Case 1. We know that f occurs in $S_2(\Gamma_0(N))$ by [E]. The claimed criterion for f to occur in $S_2(\Gamma_0(N) \cap \Gamma_1(p))$ follows from the arguments used in Case 1 of Theorem 3 on noting that if h, a newform in $S_2(\Gamma_0(N))$, is congruent to a p-new form in $S_2(\Gamma_0(N) \cap \Gamma_1(p))$, then it is in fact congruent to a p-new form in $S_2(\Gamma_0(Np))$. We may and will assume that $r \geq 2$ (i.e., ψ is not the trivial character; the other case is already dealt with in Theorem 4). We now come to the $s > 1$ subcase of Case 1. We further consider two subcases:

$\mathbf{s > 2r}$. In this case we use Theorem 2 to produce a h in the p-new part of $S_2(\Gamma_0(Np^s))$ which is congruent to f mod p. We choose a character ε of conductor p^r, of order a power of p and such that $\varepsilon^2 = \psi$ (this may be done as we are assuming that $p > 2$). Then we twist h by ε to obtain h_ε which is in the p-new part of $S_2(\Gamma_0(Np^s), \psi)$ (by Theorem 3.1 of [A-Li]) and is congruent to f mod p.

$\mathbf{r \leq s \leq 2r}$. The case $s = r$ is dealt with by Carayol's lemma. So if we have a primitive character ψ of conductor p^r $(r \leq s)$ and of order a power of p then the lemma of Carayol will imply that f occurs (necessarily in the p-new part) of $S_2(\Gamma_0(Np^r), \psi))$. Let us denote the congruent form by g. Then it is easily seen that the automorphic representation corresponding to g is principal series at p. The local

component at p of π_g is in fact of the form $\pi(\varepsilon_1, \varepsilon_2)$ where the ε_i are quasicharacters of $\mathbb{Q}_p$ with one of them being unramified (see Lemma 10.1 of [H]).

Now let us assume further that s is greater than $r+1$. Then in this case we can twist g by a character ε of conductor p^{s-r} and of order a power of p (this exists by our assumption of the previous sentence). In the case when $s = 2r$ we demand further that the character ε we twist by is such that the conductor of $\psi\varepsilon$ and $\psi\varepsilon^2$ is still p^r (this can be done by our assumption that $p \geq 5$; in fact p greater than 2 suffices here). Then the resulting form g', by what we have noted about the p-component of π_g, is easily checked to occur in the p-new part of $S_2(\Gamma_0(Np^s), \psi')$, for ψ' a character of conductor p^r and of p power order (as we are assuming that $s \leq 2r$ and in the case when $s = 2r$ we have been careful in our choice of ε). This can be seen either from what we have noted about the local component at p of π_g or Theorem 3.1 of [A-Li]. But then as all characters of conductor p^r and of p power order are conjugate under the action of the inertia group of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ (we are imagining that we have fixed an embedding of $\overline{\mathbb{Q}} \to \overline{\mathbb{Q}}_p$), we may see that f occurs mod p in the p-new part of $S_2(\Gamma_0(Np^s), \psi)$.

We now treat the case when $s = r+1$ which is treated differently to the above. It is easily seen that we need only consider the case when $r \geq 2$. In this case we consider the form g in $S_2(\Gamma_0(N))$ which is congruent mod p to f. Then we twist the form by ω to get g_ω, ω being the Teichmuller character. Then by Carayol's lemma we see that g_ω occurs mod p in (the necessarily p-new part of) $S_2(\Gamma_0(Np^r), \psi\omega^2)$ (this follows from our assumption that $r \geq 2$). Denote this congruent form by h_ω. But now it is easily seen that we can twist h_ω by ω^{-1} to obtain a eigenform in the p-new part of $S_2(\Gamma_0(Np^s), \psi)$ which is congruent to f mod p. This finishes off Case 1.

Case 2.

We deal first with the subcase $s > 2$ of this case. By [E] we note that f mod p occurs in the p-new part of $S_2(\Gamma_0(Np), \omega^{k(\rho)-2})$ (we are abbreviating $\rho_{f,p}$ to ρ and ω is the Teichmuller character whose mod p reduction had been called χ in §4). Now exactly as in the proof of Theorem 2 we may show that f occurs mod p in the

p-new part of $S_2(\Gamma_0(Np^s), \omega^{k(\rho)-2})$ for all $s \geq 3$. At this point we can now argue as we have just done in Case 1 and thus we shall be brief.

s > 2r. Twisting the p-new form congruent to f in $S_2(\Gamma_0(Np^s), \omega^{k(\rho)-2})$ by suitable character of conductor p^r, we are done in this case by another application of Theorem 3.1 of [A-Li].

r ≤ s ≤ 2r. Here again we use Carayol's lemma to produce a congruent form in (the necessarily p-new part of) $S_2(\Gamma_0(Np^r), \psi)$ which will necessarily have its local component at p to be principal series. Proceeding exactly as in the analogous subcase of Case 1, considering separately the cases s different from $r+1$ and $s = r+1$, we see that f mod p occurs in the p-new part of $S_2(\Gamma_0(Np^s), \psi')$ where ψ' is a character of conductor p^r which is congruent to ψ modulo a place above p. Then we are home by acting on this congruent form by a suitable element of the inertia group at p just as we saw earlier in Case 1.

Now we come to the subcase of Case 2 which corresponds to $s \leq 2$. We know by [E], as we have noted, that f mod p occurs in the p-new part of $S_2(\Gamma_0(Np), \omega^{k(\rho)-2})$. If the invariant $k(\rho_{f,p})$ is not $p+1$, then what we have claimed follows on noting that if g is a newform in $S_2(\Gamma_0(Np), \omega^k)$ and k is not congruent to 0 mod $p-1$, then $a_p(g)\overline{a}_p(g) = p$. Then by virtue of this fact we see, by an argument similar to the proof of Theorem 2, that f occurs mod p in the p-new part of $S_2(\Gamma_0(Np^2), \omega^{k(\rho)-2})$. Also by Carayol's lemma it occurs in (the necessarily p-new part of) $S_2(\Gamma_0(Np^2), \psi)$ for any character of conductor p^2 which is congruent to $\omega^{k(\rho)-2}$ modulo a place above p. If the invariant is $p+1$ then what is claimed in the case $s = 2$ follows from our proof of Case 3 of Theorem 3.

Case 3. By [E], [A-S] and Carayol's lemma we know that f occurs mod p in $S_2(\Gamma_0(Np^2), \omega^{k(\rho)-2})$. Any eigenform in the p-new part of this space has pth eigenvalue 0. After this what we have claimed follows from an argument which is too close to the one given in the earlier cases to bear repetition.

10.3 It may be interesting to analyse the local behaviour at p of the automorphic representations corresponding to the congruent forms we produce in Theorem 4.

The answer is likely to be a little involved as 10.1 above suggests. As we see from this that starting from a form f in $S_2(\Gamma_0(N))$ such that the mod p representation (assumed to be irreducible) is non semisimple when restricted to D_p, we produce a newform in the p-new part of $S_2(\Gamma_0(Np^2))$ which is congruent to f and supercuspidal at p. But by Remark 8 we see that we can get a p-new form in $S_2(\Gamma_1(Np^2))$ which is congruent to f mod p and is principal series at p (as we can choose the nebentype character of a congruent form to have conductor p^2). We hope to come back to the question of analysing the behaviour of the congruent forms at p in a future work.

10.4 We also remark that from Theorem 3 we may deduce that there cannot be a mod p congruence between a newform f (where as usual we assume that the corresponding mod p representation is absolutely irreducible) in $S_2(\Gamma_0(p))$ and a newform in $S_2(\Gamma_0(p^2))$. As we see from Mazur's principle in [R 1] that the mod p representation associated to f cannot be finite at p (as this would mean that $\rho_{f,p}$ arises from $S_2(SL_2(\mathbb{Z}))$ which is a contradiction). Then we are done by Case 3 of Theorem 3. In the light of Theorem 2 and Theorem 3 we may conclude that even in the (p,p) situation of this chapter, congruences are always reflected in some geometric property of the Jacobian, i.e., all congruences can be obtained by studying intersections of p-old and p-new subvarieties of the Jacobian.

10.5 We give a numerical example which illustrates our results which we take from the tables of Cremona, cf. [C]. We see that the arguments in Case 2 of Theorem 3 (these are valid for $p = 2, 3$) predict congruences between the unique form of level 11 and some newform of level divisible by 4 (respectively 9) and dividing 44 (respectively 99) mod 2 (respectively 3). We note that it is known that the mod ℓ Galois representation associated to the elliptic curve 11A is absolutely irreducible for all primes ℓ not equal to 5 (see page 309 of [Se 1]). We check that the forms associated to the elliptic curves 11A and 44A are congruent mod 2 (i.e., from the tables in [Cr], we verify that their eigenvalues at Hecke operators T_r for all primes ≤ 97 and not equal to 2 are congruent mod 2). Looking at the table we also check that the forms associated to the elliptic curves 11A and 99C are congruent mod 3

(i.e., their eigenvalues for T_r for $r \leq 99$ and not 3 are congruent mod 3). Similarly for $p = 5$, we check that the forms associated to the elliptic curves 14A and 350E are congruent mod 5. The nature of the illustrative examples is of necessity selective because the table we are consulting lists only newforms (of weight 2) with rational Fourier coefficients.

1.6. Mainly multiplicities

In this section when we speak of the Hecke algebra we shall mean the full Hecke algebra, i.e., with all the Hecke operators including T_p.

We note that if m whose residue field is of characteristic p, is the maximal ideal in the Hecke algebra $\mathbb{T}_{Np^r}$ associated to f, or f'' (this is the p-old form associated to f of §1.3), and V is the associated irreducible Galois module of dimension 2 over the residue field of m, then using the map α, we can embed different copies of V into the intersection of the p-old and p-new subvarieties as above. In our proof of Theorem 2 we chose a particular embedding, i.e., one of the form $(V, 0)$. But the diagonal and anti-diagonal embeddings of V in $J_0(Np^r)^2$ are mapped injectively into the intersection too. The fact that they are mapped to the intersection follows from the considerations in the proof of Theorem 2 when one takes into account the relation ${T_p}^*.{\alpha_1}^* = p.{\alpha_p}^*$ and the fact that if T_p induces the zero endomorphism on the abelian subvariety A_f of $J_0(Np^r)$ attached to a newform f for $\Gamma_0(Np^r)$, then ${T_p}^*$ also induces the zero endomorphism of A_f. To see that they are embedded, we will treat only the case of V embedded diagonally the other case being similar. If (x, x), for $x \in V$ is in the kernel of α, then under our assumption that V is irreducible as a Galois module, Theorem 1 shows that $(x, x) = (-{\alpha_p}^*(x_1), {\alpha_1}^*(x_1))$, for $x_1 \in J_0(Np^{r-1})$. But we note that from this relation $\alpha(x_1, x_1) = 0$. Assuming $r \geq 2$, we use Theorem 1 again to conclude that $(x_1, x_1) = (-{\alpha_p}^*(x_2), {\alpha_1}^*(x_2))$, for $x_2 \in J_0(Np^{r-2})$. We continue in this way to produce a sequence of points $x_1, \cdots, x_r$, where $x_i \in J_0(Np^{r-i})$, $\alpha(x_i, x_i) = 0$ and $(x_{i-1}, x_{i-1}) = \beta(x_i)$, to revert to notation of §1.2. Now we can use the result in [R] to get a contradiction as it is proved in [R] that the kernel of the map $\alpha : J_0(N)^2 \to J_0(Np)$ is Eisenstein. This contradicts our

assumption that V is absolutely irreducible. The same argument shows that if X is a Hecke stable submodule of $J_0(Np^r)$ which has no maximal ideals of the Hecke algebra which are Eisenstein in its support, then α is injective on the diagonal and anti-diagonal copy of X in $J_0(Np^r)^2$.

By a further elaboration of these ideas one may in fact show that the natural map $\alpha : V \times V \to J_0(Np^{r+1})$ is injective. To see this we argue by contradiction and pick the minimal s such that a two-dimensional Galois representation isomorphic to V occurs in $J_0(Np^s)$ and such that the map $V \times V \to J_0(Np^{s+1})$ is not injective. The case when $s = 0$ being the theorem of Ribet that we have already quoted, cf. [R], we may and do assume that $s \geq 1$. But if we have that ${\alpha_1}^*(x) + {\alpha_p}^*(y) = 0$ for $x, y \in V$, then either we contradict the injectivity of ${\alpha_1}^*$ and ${\alpha_p}^*$ or by using Theorem 1 we see that $y = {\alpha_1}^*(z)$, for non-zero $z \in J_0(Np^{s-1})$, which contradicts the minimality of s, upon once again using the absolute irreducibility of V as a Galois module, as this shows that V arises from $J_0(Np^{s-1})$ by pull back by α_1. (We note that this paragraph does not render redundant the preceding paragraph because, as we have already noted, in the argument in this paragraph the assumption that V is irreducible is crucial while to prove injectivity for α restricted to the diagonal and anti-diagonal embedding of a Hecke module X in $J_0(Np^r)^2$ we only need to assume that there are no Eisenstein ideals in the support of X (so for instance it could be the sum of several 2-dimensional irreducible representations such as V).)

The same argument (essentially) shows that if V is an absolutely irreducible two-dimensional Galois representation associated to $J_0(Np^r)[m]$, then the natural map

$$V \times \cdots V \to J_0(Np^{r+t}) \qquad (*)$$

from $t+1$ copies of V to $J_0(Np^{r+t})$ is injective. We argue as before but use Theorem $1'$ which shows that any (non-Eisenstein) element of the kernel of the sum of the natural degeneracy maps ${\alpha_1}^* + \cdots + {\alpha_{p^t}}^*$ from

$$J_0(Np^r)^{t+1} \to J_0(Np^{r+t})$$

is of the form $(-, \cdots, {\alpha_1}^*(x))$ where $x \in J_0(Np^{r-1})$ and where we are continuing with our abuse of notation in conflating degeneracy maps arising from varying p

power levels. Thus we use induction on t, the case $t = 1$ being already handled. We pick a minimal s, which we suppose to be greater than 0, such that a Galois module isomorphic to V occurs in $J_0(Np^s)$ and the map corresponding to $(*)$ is not injective. Then by what we have noted about the shape of the kernel of the degeneracy map in this situation, we see either that we contradict the minimality of s or we are reduced to the case $t-1$. Here again our hypothesis that V is irreducible is crucial as we are using the fact that if V intersects non-trivially the image of the pull back of the Jacobian of a lower level then it in fact lies entirely in the image of the pull back. The base case when $s = 0$ is handled by a straightforward generalisation of Ribet's theorem which will say that the natural map

$$J_0(N)^{t+1} \to J_0(Np^t)$$

has Eisenstein kernel. This shows that two-dimensional Galois representations occur with high multiplicity in $J_0(Np^{r+t})$. We state this in the form of a proposition.

Proposition 2. *If m is a maximal ideal of $\mathbb{T}_{Np^{r+t}}$, as in the discussion above (and hence of residual characteristic p), then if m arises by pull back from $\mathbb{T}_{Np^r}$ from an ideal which is new of level divisible by p^r, $r > 1$, the multiplicity of $J_0(Np^{r+t})[m]$ is at least $t + 1$.*

Proof. From the above discussion it is enough to note that T_p annihilates the image of $V \times \cdots \times V$ in $J_0(Np^{r+t})$. This follows from our assumption that $r > 1$ and the relation $T_p.{\alpha_p}^* = p.{\alpha_1}^*$.

Remark 11. We can give a more appealing form to the above proposition. As what we can say from the above discussion is that $J_0(Np^\infty)[m]$, for m any maximal ideal of characteristic p of the Hecke algebra $\mathbb{T}_{Np^\infty}$ which contains T_p (i.e., is non-ordinary), is infinite dimensional as a $\mathbb{T}_{Np^\infty}/m$ vector space. Here by $\mathbb{T}_{Np^\infty}$ we just mean the inverse limit of $\mathbb{T}_{Np^r}$, for $r \geq 1$, taken with respect to the natural maps which send T_n to T_n and by $J_0(Np^\infty)$ we mean the direct limit of $J_0(Np^r)$, $r \geq 1$, induced by the maps ${\alpha_1}^*$. We may naturally consider $J_0(Np^\infty)$ as a $\mathbb{T}_{Np^\infty}$ module.

Remark 12. Much work has been done on the question of multiplicities, but for a study of this question which is close to the present situation see [M-R]. In [M-R]

examples of higher multiplicity are shown to occur in $J_0(Np^3)$. But one may in fact see that multiplicity 1 already fails for $J_0(Np^2)$ by the above methods. For this one need only look at a maximal ideal m of $\mathbb{T}_{Np}$ which is not ordinary at p (such a m is bound to exist if we assume that $X_0(N)$ has genus greater than 0). Then if we assume that ρ_m is irreducible and consider the associated Galois representation V which is 2 dimensional over $\mathbb{T}_{Np}/m$ and which occurs in $J_0(Np)$ then the image of $V \times V$ under the degeneracy map $J_0(Np)^2 \to J_0(Np^2)$ furnishes examples of higher multiplicity occurring in $J_0(Np^2)$. As, by the above, the degeneracy map is injective on $V \times V$ and all the Hecke operators (including T_p) act diagonally on the image of $V \times V$, the Hecke operators away from p being equivariant with respect to α (T_p in its turn annihilates the image of $V \times V$ under α). In contrast to the proposition, it may be noted that in the case when the residual characteristic of a maximal ideal of a Hecke algebra is prime to 2 times the level, then the corresponding multiplicity is 1. This is proved in [R 1] using the techniques of [M 1]. We may also note that we do not know if multiplicity 1 holds for the p-new quotient of $J_0(Np^r)$. The proposition does not rule this out although as a consequence of the above discussion one cannot expect multiplicity 1 for the p-new subvariety.

Remark 13. If we consider the natural degeneracy map:

$$\alpha : J_0(Np^r)^{n+1} \to J_0(Np^{r+n}) \qquad (*)$$

(see Remark 3 of §1.2) then the composition of the dual map:

$$\alpha^\vee : J_0(Np^{r+n}) \to J_0(Np^r)^{n+1}$$

with α when written as a $(n+1) \times (n+1)$ matrix of endomorphisms of $J_0(Np^r)$ (we shall denote this by R_n) has (i,j)th entry given by ${\alpha_{p^i}}_* . {\alpha_{p^j}}^*$ and for $j < i$ this is ${T_p}^{i-j}$. Further we note that the diagonal entries of the matrix are p^n as these are the degrees of the map α_{p^i}'s. From this we see that if A is an abelian subvariety of $J_0(Np^r)$ on which T_p acts as the zero endomorphism, then $A[p^n]$ when regarded as embedded in $J_0(Np^r)^{n+1}$ via $a \to (a, \cdots, 0)$ (with all entries except the first equal to 0), is in the kernel of R_n. Also by inspection we see that α is injective on this

copy of $A[p^n]$. We have seen in Theorem 1′ that the group of connected components of the kernel of α and hence that of the kernel of $\alpha^\vee$ is Eisenstein. Thus we see that α embeds $A[p^n]$ in the connected component of the identity of the kernel of $\alpha^\vee$ if we assume that the ideals in the support of $A[p^n]$ are non-Eisenstein. The space of regular differentials of this abelian variety is spannned by forms whose minimal p power level is greater than r. We may state this qualitatively in the form of the following proposition.

Proposition 3. *We assume $r \geq 2$. Then the intersection of the sum of the images of $J_0(Np^r)$ under the degeneracy maps in $J_0(Np^\infty)$ with the direct limit of abelian subvarieties of $J_0(Np^\infty)$ whose space of differentials is spanned by forms whose minimal p power level is divisible by p^{r+1} has infinite p exponent, i.e., is not annihilated by any fixed power of p (assuming that there is a maximal ideal in the support of $J_0(Np^r)^{p-new}$ which is non-Eisenstein; this is generally the case).*

1.7. Modular deformations

In this section we assume that $p \geq 5$. Mazur, in [M], has defined the universal ring of deformations of an irreducible mod p representation of the Galois group of $\mathbb{Q}$. If the mod p representation is itself modular, then it is of interest to study the locus in the space of all deformations which corresponds to modular deformations. In [M] a lower bound on the Krull dimension of the ring of deformations is given. We would like to give a lower bound on the ring of modular deformations. There is a discrepancy in the bounds. We are not sure if it arises from the limitations of our proof, or represents a genuine gap between all deformations and modular deformations. This lower bound has been proven earlier and is Proposition III.6.13 of [G]. In [G] information of the p-adic Hodge structure of the p-adic representations corresponding to classical eigenforms is used. We present a different, more modular approach. We have two proofs of this which are rather different in spirit. One of the proofs relies on weight variation, and relies upon results of Serre, Tate and Jochnowitz, cf. [J], while the other, which we give here, relies on level variation, and uses Theorem 2. The ring of modular deformations is identified with a local

component of a certain universal Hecke algebra, whose definition we will recall presently. For this we need to set up some notation.

We denote by h_r, the $\mathbb{Z}_p$ algebra generated by the image of the Hecke operators T_n in the ring of endomorphisms of the space of cusp forms of weight 2 for the group $\Gamma_1(Np^r)$ with coefficients in $\mathbb{Q}_p/\mathbb{Z}_p$ which we denote by $S_2(\Gamma_1(Np^r), \mathbb{Q}_p/\mathbb{Z}_p)$. Here we suppose that $(N,p) = 1$ and $r > 1$. Then we may take the inverse limit of the h_r's with respect to the maps which are induced on them by the natural inclusion maps:

$$S_2(\Gamma_1(Np^r), \mathbb{Q}_p/\mathbb{Z}_p) \quad \hookrightarrow S_2(\Gamma_1(Np^s), \mathbb{Q}_p/\mathbb{Z}_p)$$

for $s \geq r$. Then we define the universal Hecke algebra in our situation by:

$$h^{\text{univ}} = \varprojlim h_r.$$

h^{univ} is a semi-local ring and we write it in terms of its local components as

$$h^{\text{univ}} = \prod_i R_i$$

where the R_i are complete local rings. It is known that h^{univ} has only finitely many components which is equivalent to the fact that there are only finitely many Hecke eigensystems of weight 2 and level Np^∞ mod p (see Remark 6 of §1.2). We shall focus our attention on a given local component, say R, and denote the corresponding maximal ideal by m. We further assume that the corresponding mod p representation is absolutely irreducible. Then it follows from Mazur's theory of deformations in [M] that R is Noetherian. This is shown in [H 1] and we refer to that for a discussion of these issues. We briefly recall the relevant part of the discussion in [H 1]. R (deprived of T_p) is the ring of modular deformations of the corresponding mod p representation which R gives rise to. In [H 1] it is shown (based on Wiles' theory of pseudo-representations) how to attach a representation of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ into $GL_2(R)$ which interpolates the representations of Eichler-Shimura attached to homomorphisms of R into $\overline{\mathbb{Q}}$ which correspond to classical Hecke eigenforms occuring in $S_2(\Gamma_0(Np^r))$ for varying r. From the properties of this representation proved in

[H 1], it follows that the deformation ring of Mazur surjects onto $R^{(Np)}$ by which we mean R deprived of the Hecke operators T_n for n not coprime to Np. But the deformation ring of Mazur is Noetherian and R is finitely generated over $R^{(Np)}$. Thus R is Noetherian. We now assume that R is a non-ordinary local component as there is a very well developed theory of ordinary components due to Hida, cf. [H]. By R being non-ordinary we mean that T_p is in the maximal ideal m of R. Now we can state the following theorem.

Theorem 5. *The Krull dimension of R is greater than or equal to 4.*

Proof. We need to recall that there is a perfect pairing between the Hecke algebra and the space of cusp forms. For this we refer to [H], and just state the result. We define the pairing,

$$h^{\text{univ}} \times S \to \mathbb{Q}_p/\mathbb{Z}_p$$

by the usual formula:

$$(h, f) = a_1(f|h). \tag{1}$$

Here S is by definition $\bigcup_{r=1}^{\infty} S_2(\Gamma_1(Np^r), \mathbb{Q}_p/\mathbb{Z}_p)$. Then in [H] it is proven that this gives a perfect pairing, i.e., S and h^{univ} are Pontryagin duals of each other. We have a natural action of the one units of ${\mathbb{Z}_p}^*$ on S given by the usual diamond action. By definition we see that the fixed part of S under this action contains:

$$S_0 := \bigcup_{r=1}^{\infty} S_2(\Gamma_1(Np) \cap \Gamma_0(Np^r), \mathbb{Q}_p/\mathbb{Z}_p).$$

(One may in fact deduce from results of Katz [Ka] that this inclusion is actually an equality: but we do not need this fact. We thank H. Hida for pointing out that this equality we had blithely asserted in an earlier version is non-trivial.) We denote the fixed space under Γ of S by S' and the group of one units of ${\mathbb{Z}_p}^*$ by Γ and choose a topological generator for it, say γ. Then by a standard result we see that the pairing (1), induces a perfect pairing:

$$R \,/\, (T_p, p, \gamma - 1) \times \widetilde{S'}_R[T_p] \to \mathbb{F}_p \tag{2}$$

where the tilda sign denotes the kernel of multiplication by p and the subscript R refers to that part of the space of cusp forms on which m is topologically nilpotent.

We can see that the image of the sequence $(T_p, \gamma - 1, p)$ in each local component is regular. We may see this by firstly noting that by the arguments in [J], T_p is a transcendental element in each non-ordinary local component (in fact by Theorem 6.3 of [J] one may see that R is a power series ring in the variable T_p over $R^{(p)}$). Secondly $\gamma - 1$ is not a zero divisor in $R\ /\ (T_p)$, as if for some $T \in R\ /\ (T_p)$ we have that $(\gamma - 1).T = 0$, then if T is not the zero endomorphism in $R\ /\ (T_p)$, there exists a g, with $g|T_p = 0$ and such that $g|T$ is not 0. Then by twisting g by a character χ of ${\mathbb{Z}_p}^*$ of sufficiently high p power order (we denote the resulting form by g_χ), we can ensure that $g_\chi|T|(\gamma - 1)$ is not zero using the well known commutation relations between twisting and the Hecke action, cf. [G]. (We note that g_χ is again in $S[T_p]$.) Then the fact that $(T_p, \gamma - 1, p)$ is a regular sequence in each local component follows from the fact that Γ is a pro-p group as by twisting by characters with values in $\mathbb{Q}_p/\mathbb{Z}_p$, we see that for any local component R, the image of the corresponding maximal ideal m in the ring of endomorphisms of $\tilde{S_0}$ is not the unit ideal (here we are using the assumption that p is not 2).

Now we may use Theorem 2 which shows that a cusp form f can be propagated mod p to arbitrarily large p power levels. More precisely, under the assumption that the local component R is non-ordinary we see that if f is an eigenform associated to R occuring mod p in $S_2(\Gamma_0(Np^r) \cap \Gamma_1(p))$ for some r, then the theorem applies, as the proof works even with the slightly different level condition on f we have, and we can increase the p power level of the form f arbitrarily mod p. This is because the proof of Theorem 2 shows that any eigencuspform f of weight 2, which is killed by T_p, has the property that it is congruent to newforms with arbitrarily large p power levels, i.e., the dimension of the subspace $S_0(\Gamma_0(Np^s) \cap \Gamma_1(p))[T_p]$ on which m is nilpotent tends to infinity with s. In other words the space $\widetilde{(S_0)}_R[T_p]$ is an infinite dimensional $\mathbb{F}_p$ algebra. We see this upon using Theorem 2 and the fact that if Θ is the operator:

$$\Theta(\sum a_n q^n) = \sum_{(n,p)=1} a_n q^n$$

then Θ maps $S_2(\Gamma_0(Np^r) \cap \Gamma_1(p))$ to itself for $r > 1$ (and as R is non-ordinary preserves the R part of the space), its image is the kernel of T_p and the kernel of

Θ on $S_2(\Gamma_0(Np^r) \cap \Gamma_1(p))$ is given by the image of $S_2(\Gamma_0(Np^{r-1}) \cap \Gamma_1(p))$ under V where V is defined by:

$$V(\sum a_n q^n) = \sum a_n q^{np}.$$

These are well known facts which may be deduced from [A-L]. Then on using perfect pairing (2) and the fact recalled above that R is Noetherian, we see that R has Krull dimension at least 4 as we have shown that $(T_p, \gamma - 1, p)$ is a regular sequence in R and that $R/(T_p, \gamma - 1, p)$ is infinite dimensional as a $\mathbb{F}_p$ vector space and hence as it is Noetherian has Krull dimension at least 1.

Remark 14. We note that by an argument similar to the above we can easily see that any ordinary local component has Krull dimension ≤ 2. This follows upon using the facts that a newform f for the group $\Gamma_1(Np^r)$, $r > 1$, with nebentype whose conductor is Np^s, for $s < r$, is annihilated by T_p and the result of Katz we alluded to above. The fact that the Krull dimension of such a local component is 2 is harder and follows from Hida's theorem that any such ordinary local component is finite and flat as a module over $\mathbb{Z}_p[[X]]$, cf. [H].

Remark 15. We have reason to believe that this lower bound is in fact the dimension of non-ordinary local components of the Hecke algebra. In the unobstructed case of the deformation problem it is proven in [M] that the dimension of the universal ring is 4. This maps surjectively to the corresponding local component of our Hecke algebra when it has been deprived of the Hecke operators T_n for n not comprime to Np. But it is unlikely that all deformations are modular, as we are considering unrestricted deformations. Thus as in the unobstructed case the universal ring is a power series ring, it seems likely that the dimension of the corresponding T_p deprived local component of the Hecke algebra will be at least, and hence by Theorem 5 exactly, 1 less than the dimension of the universal ring.

Chapter 2

Mod p modular forms

2.1. Introduction

In this chapter we study the space of mod p modular forms of Serre and Swinnerton-Dyer especially with a view to understanding the structure of the mod p Hecke algebra. We rely upon the paper of Jochnowitz, cf. [J], heavily to obtain some information about local components of the mod p Hecke algebra. The new ingredient is the fact that it is now known from the deformation theory of Mazur, cf. [M], that the local components are Noetherian. This allows one to strengthen some of the results of [J]. We also include in the last section of this chapter speculation about certain sequences of Hecke modules which we believe hold the key to understanding local components better.

2.2. Varying the weight

We fix a congruence subgroup Γ of $SL_2(\mathbb{Z})$ and for any integer $k \geq 2$ denote by $A_k(\Gamma)$ the space of weight k modular forms for Γ such that the coefficients of the Fourier expansion at infinity are in $\overline{\mathbb{Q}}$ and are integral for the valuation induced by a choice of a prime ideal above p. We denote by M_k the image of A_k in $\overline{\mathbb{Q}}[\,[\,q\,]\,]$ where hereafter we eliminate the mention of Γ as we will consider modular forms only for this fixed Γ. We denote by $\widetilde{M_k}$ the reduction modulo the chosen prime ideal lying above p. Then we have inclusions $\widetilde{M_k} \subset \widetilde{M_{k+p-1}}$. We denote by $\widetilde{M}^a$, where $a \epsilon \mathbb{Z}/(p-1)$, the union of the $\widetilde{M_k}$'s for $k \geq 2$ and $k \equiv a \bmod (p-1)$. On this space we may define a filtration w by defining $w(f)$ to be the least k such that f is in the image of M_k under the reduction map. As usual we have the Hecke action. We only note here that the action of the p th Hecke operator, which we denote by U, is

$$\sum a_n q^n \to \sum a_{np} q^n.$$

We denote by $\widetilde{M}$ the space $\bigoplus_a \widetilde{M}^a$. This is for us the space of mod p modular forms for the group Γ which comes equipped with a grading mod $p-1$. On the space of mod p forms one has several interesting operators, chief amongst these being the Θ operator and V which are defined by their effect on q expansions by :

$$\Theta(\sum a_n q^n) = \sum n a_n q^n$$

and

$$V(\sum a_n q^n) = \sum a_n q^{np}.$$

These preserve the space of mod p modular forms. Their effect on the filtration is given by $w(\theta(f)) \leq w(f) + p + 1$ and $w(V(f)) = pw(f)$. These results follow from [Ka 1]. It is the existence of these operators which makes the theory of mod p modular forms more tractable than that in characteristic 0, as was discovered by Serre, Swinnerton-Dyer and Tate.

We note the basic relations between the operators U, V and Θ considered as operators on the space of mod p modular forms. $ker(\Theta) = Im(V)$, $ker(U) = Im(\Theta)$, V is injective and U is surjective. These facts follow from the definitions of these operators on noting that UV is the identity map.

Note that any $f \epsilon \widetilde{M}$ has the decomposition:

$$f = \sum f_i$$

where $f_i \epsilon \overline{\mathbb{F}}_p[\,[\,q\,]\,]$ is such that only exponents of q of the form q^{np^i}, where $(n,p) = 1$, occur with non-zero coefficient. As f_i is checked to be given by

$$f_i = \Theta^{p-1}(f|U^n)$$

we see that it is a mod p modular form. The filtration of f_i satisfies, $w(f_i) \leq w(f) + p^2 - 1$. So in particular the filtrations of the f_i 's are bounded. This

suggests that the image of $\widetilde{M}$ inside $M = \widehat{\bigoplus} V^n \Theta^{p-1}(\widetilde{M})$ is small. Here $\widehat{\bigoplus}$ denotes unrestricted direct sum. We note that the operator Θ^{p-1} is idempotent. We denote it by Φ. We remark that if p does not divide $w(f)$, then we have $w(f|\Phi) \geq w(f)$. We note that on $\bigoplus V^n \Phi(\widetilde{M})$ the operator U is nilpotent. Using the lemma of Jochnowitz, $w(f|U) < w(f)$ if $w(f) > p+1$, we show in the following Proposition that the image of $\widetilde{M}$ in M is rather small.

Proposition 1. *$M/\bigoplus V^n \Phi(\widetilde{M})$ is a finite dimensional $\overline{\mathbb{F}}_p$ vector space.*

Proof. We denote the vector space in the Proposition by S. We note that U acts bijectively on S. Note that by the lemma of Jochnowitz that we have quoted, the dimension of a finite dimensional subspace of $\widetilde{M}$ on which U is injective has bounded dimension. This implies that the space S is finite dimensional.

We rewrite the proof of the lemma of Jochnowitz.

Write $f = \Phi(f) + g|V$. Note that $f|U = g$. We also have that $w(g) \leq \frac{w(f)+p^2-1}{p}$. From this the result follows.

Remark. The lemma of Jochnowitz and the above proposition occur already in the article of Serre [Se 4].

In the light of the above Proposition, understanding modular forms mod p is reduced to studying $\Phi(\widetilde{M})$. Note that one may also characterise this vector space as the kernel of U acting on $\widetilde{M}$.

We may deduce some amusing consequences from the Proposition. For example, it immediately implies that there exists a r, such that any $f \in \widetilde{M}$ can be written as

$$f = g|\frac{P(V)}{\sum_{i=0}^{r} a_i V^i}$$

where $g \in \Phi(\widetilde{M})$, P is a polynomial and $a_i \in \overline{\mathbb{F}}_p$. Thus if $f|R(V)$ is an element of $\widetilde{M}$, for $f \in \Phi(\widetilde{M})$, then $R(V)$ is a rational function with denominator of bounded degree. Here we remark that the non-zero elements of $\overline{\mathbb{F}_p}\,[\,[\,V\,]\,]$ act injectively on $\overline{\mathbb{F}_p}\,[\,[\,q\,]\,]$, so all this makes sense!

Now we study local components of the mod p Hecke algebra. We set up some notation. The inclusion

$$\widetilde{M_k} \subset \widetilde{M_{k+p-1}}$$

induces a map of the Hecke algebras $h_{k+p-1} \to h_k$ where h_r in this chapter is the image of the Hecke operators in the endomorphisms of $\widetilde{M_r}$. We define the limit Hecke algebra H as the algebra obtained as the inverse limit of the h_r's taken with respect to these maps. H is a semi-local ring with finitely many local components, cf. [J]. We focus on a local component R and denote its maximal ideal by m_R and assume that U belongs to m_R as in the other ordinary case it follows by Proposition 1 that the local component is uninteresting. In [J] it is proven that local components for which the operator U is nilpotent, have Zariski tangent space dimension bigger than 2. In fact using her arguments, which prove that any local component R on which U is nilpotent has the property that $R/(U)$ is infinite dimensional over $\overline{\mathbb{F}_p}$, and the fact which is known by Mazur's theory of deformations that any local component is Noetherian, we see that all local components for which U is nilpotent, have Krull dimension ≥ 2. We now give an argument for this after stating it as a proposition.

Proposition 2. *If R is a non ordinary local component of H, then R is a power series ring in U over $R^{(p)}$ where by this we mean the subalgebra otf H topologically generated by operators T_n for $(n,p) = 1$. Further, if we assume that $p \geq 3$ and that the mod p representation attached to R is absolutely irreducible, then the Krull dimension of R is at least 2.*

Proof. The first part is proven as Theorem 6.3 of [J] but we recall the main point of the proof. We need to check that no expression of the form $\sum c_i U^i$ is 0 in R where the c_i's belong to $R^{(p)}$. We assume that n is the least i for which c_i is not 0. Then there is a g in $\widetilde{M}$ on which m_R acts nilpotently and such that c_n does not kill g. We may assume that $g|\Theta$ is not 0 by applying an appropriate power of U to g. Then $g|\Phi|V^n$ is not in the kernel of U^n but is in the kernel of U^m for all $m > n$. This shows what we wanted.

Now we prove that the Krull dimension of R is at least 2 under the assump-

tions in the proposition. We denote by $\widetilde{M_R}$ the subspace of $\widetilde{M}$ on which m_R is locally nilpotent. By Corollary 6.6 of [J] it follows that $\widetilde{M_R}/V(\widetilde{M_R})$ is an infinite dimensional vector space over $\overline{\mathbb{F}}_p$. From this it follows upon using the perfect pairing

$$R \times \widetilde{M_R} \to \overline{\mathbb{F}}_p$$

that $R^{(p)}$ is infinite dimensional as a $\overline{\mathbb{F}}_p$ algebra. Now we use the fact, known from [M], that R and hence $R^{(p)}$ is Noetherian. This follows from [M] on noting that the universal ring corresponding to the mod p representation associated to R is Noetherian and surjects onto $R^{(Np)}$ where N is the level of Γ and as usual by this notation we mean R deprived of operators T_n which are not coprime to Np. This surjection is a consequence of the construction in [H 1] of a big representation into $GL_2(\widetilde{R})$ which deforms the corresponding mod p representation and where $\widetilde{R}$ is the chracteristic 0 Hecke algebra which surjects onto R. As $\widetilde{R^{(Np)}}$ is identified as the $\mathbb{Z}_p$ algebra generated by the traces of the Frobenii for primes outside Np (at which this representation is unramified) by the properties of this big modular deformation we check easily that the deformation ring of Mazur surjects onto $\widetilde{R^{(Np)}}$. (The situation in [H 1] is slightly different but the method of proof which uses the pseudo-representations of Wiles is easily checked to work in our situation as we are assuming that $p \geq 3$ and that the mod p representation is absolutely irreducible.) For the precise definition of $\widetilde{R}$ (in the context of cusp forms) we refer to the later part of this section. Thus we know that $R^{(p)}$ has Krull dimension at least 1 and hence that R has Krull dimension at least 2.

Its also proven in [J] that most local components have Zariski tangent space dimension ≥ 3. It is hoped that the Krull dimension of local components is bounded by 2 for reasons we shall go into later. Thus this indicates that the non-ordinary local components, as we shall henceforth call the local components for which U is nilpotent, are not regular.

We now suggest a method for proving that the Krull dimension of local components of the Hecke algebra are bounded by 2. For this we need only study the space

$$\bigoplus V^n \Phi(\widetilde{M})$$

as this is the subspace on which U is locally nilpotent. In order to study one of these local components, R, it will be enough to study $R/(U)$ by Proposition 2. This acts naturally on the part of $\widetilde{M}$ which is killed by U. There is a perfect pairing between $R/(U)$ and the subspace of M on which the maximal ideal of R is locally nilpotent intersected with the kernel of U. For this we may again refer to [J]. The pairing, which is given by $(T, f) = a_1(f|T)$, has the property that the Hecke action with respect to this pairing is self adjoint. So as notation we denote $R/(U)$ by $\mathcal{R}$, and the above space which it is in duality with, by $\mathcal{M}_R$.

Denote the kernel of U, viewed acting on $\widetilde{M}$, by $\mathcal{M}$. We may consider $\mathcal{M}$ as a module for the Hecke algebra H quotiented by U. We see, using the fact that all eigenforms have filtration $\leq p^2 - 1$, proven in Ash and Stevens [A-S], that $\widetilde{M^a}$ is an essential extension of $\widetilde{M_k}$, where $a \equiv k \pmod{p-1}$, and $k > p^2 - 1$, as H modules. Here we use the terminology in the appendix of the book of Matsumura, cf. [Ma]. This is a direct consequence of the result of Ash and Stevens, and the fact that an eigenvector of all the Hecke operators T_n is determined upto scalar by knowing all its eigenvalues. Everything above also holds when we only look at the kernel of U and consider the resulting exact sequences as $H/(U)$ modules. In order to prove bounds on the Krull dimensions of local components it will be enough to find a Hecke operator T_n, such that the above statements are true as $\overline{\mathbb{F}}_p[T_n]$ modules for sufficiently large k. We expand on this remark.

What we see is that for sufficiently large k, $\mathcal{M}_R$ is an essential extension of $(\mathcal{M}_R)_k$ as $\mathcal{R}$ module. If we could show an analogous statement for the subspaces in the kernel of U for the subalgebra generated by a single element say a Hecke operator T_n, for some n, then we could bound the Krull dimension by linear algebra (Jordan canonical form) as this will show that $(\mathcal{M}_R)_k$ is of bounded rank when considered as a $\overline{\mathbb{F}}_p[T_n]$ module. This amounts to showing that the kernel of $T_n - \lambda$, acting on $\mathcal{M}_R$, where $T_n - \lambda \in \mathrm{m_R}$, is finite dimensional. This of course seems rather hard but at least gives scope for computation. For example one may try to compute all the eigenvectors for T_2 acting on mod 3 forms of all weights ≥ 2 and level 1. This is just the polynomial algebra in Δ. There is only one local component for which

U is nilpotent. Calculations have been done by Prof. Maeda in this case (whom we thank heartily). The computations are a little inconclusive though they do show that the eigenvectors of T_2 are rather sparse.

On a theoretical level, the above approach leads to some interesting questions. One of the distinctive features of mod p modular forms is that there are only finitely many eigensystems. The finiteness of the number of eigensystems comes from the fact, due to Serre and Tate, that the quotients $\widetilde{M_{k+p-1}}/\widetilde{M_k}$ are only finitely many, upto isomorphism, when considered as Hecke modules. From this one deduces the fact that

$$0 \to \widetilde{M_k} \to \widetilde{M_{k+p-1}} \to SS_k \to 0$$

is an essential extension of Hecke modules for large enough k. Here SS_k is defined to make the above sequence exact as Hecke modules. As remarked above, this is seen on combining the above mentioned result of Serre and Tate, with the fact that an eigenform for all the Hecke operators is determined upto scalar by its eigenvalues. In a possible approach to get hold of an operator T_n with the properties described above, and which we describe slightly later, the following question comes up :

Given a integer $r \geq 2$, is it true that for sufficiently large k in the congruence class of r mod $p-1$, the following exact sequence is an essential extension of Hecke modules:

$$0 \to \widetilde{M_k}/\widetilde{M_r} \to \widetilde{M_{k+p-1}}/\widetilde{M_r} \to SS_k \to 0 \quad ?$$

The above method of proof, when $\widetilde{M_r}$ was taken to be zero, will not work as we no longer have the fact that eigenforms are determined by their eigenvalues upto scalars in the space $\widetilde{M}/\widetilde{M_r}$. But we can answer this question affirmatively in most cases (see Proposition 4 of §2.3).

Another issue which comes up is to determine the structure of $\mathcal{M}_R$ as a $\mathcal{R}$ module. Denoting the maximal ideal of $\mathcal{R}$ as above by m_R, we see that $\mathcal{M}_R[\mathrm{m}_R]$

is a $\mathcal{R}/\mathfrak{m}_R$ module of rank 1. If we knew the same about $\mathcal{M}_R/\mathfrak{m}_R\mathcal{M}_R$, then we would be home by Nakayama's lemma. But we do not know how to prove this.

The tentative approach we would suggest to the question of determining Krull dimensions of local components of Hecke algebras is the following:

We will first try to prove a much stronger form of the above. Thus what we will need is that we should be able to answer the above type of question with an explicit bound on the k which will work, in terms of r (see Conjecture 2 of the next section). This question is probing the relationship between the two natural filtrations on the space of modular forms, i.e., one coming from the weight and one coming from the degree of nilpotence with respect to the maximal ideal of the local component. We see from the fact that the Krull dimensions of local components are at least 2 that the relationship between the two may not be straightforward for $\widetilde{M}$. There may be a relationship between the filtrations when one passes to the kernel of U, i.e., $\mathcal{M}$. (We study this more in the §2.4 of this chapter.) Then we will have to prove a version of the theorem of Serre and Tate to say that there are only finitely many isomorphism types amongst the Hecke modules $(\mathcal{M})_{k+r(p-1)}/(\mathcal{M})_k$, for a fixed r, such that the isomorphisms respect the filtrations. Then for any given p, it may be possible to find the T_n of our dreams by a finite amount of computation.

We now prove something concretely about local components of the characteristic 0 Hecke algebra. In Theorem 5 of Chapter 1 we gave a lower bound for local components which corresponded to absolutely irreducible mod p representations in the Hecke algebra which was made by varying the p power level. But in the optic of this chapter we shall give a similar lower bound on local components of Hecke algebras made with varying the weight. It is theorem of Shimura that these two Hecke algebras, i.e., one made with varying level and the other with varying weight, are isomorphic via an isomorphism which takes T_n to T_n, cf. [H]. But perhaps it is good to give a proof in the spirit of this section. The appeal to Theorem 1 in the proof of Theorem 4 of Chapter 1 will be replaced by an appeal to Proposition 2 of this chapter in the discussion below. A background reference for what is to follow is [H 3]. We base our discussion on that.

We fix an integer N prime to p and assume that $p \geq 5$ (for safety!). We choose a finite extension K of $\mathbb{Q}_p$ and define the space of cusp forms for $\Gamma_1(N)$ with coefficients in K by:

$$S_k(\Gamma_1(N), K) := S_k(\Gamma_1(N), \mathbb{Q}) \otimes K$$

where we take the tensor product over $\mathbb{Q}$ and where as usual by $S_k(\Gamma_1(N), \mathbb{Q})$ we mean the space of cusp forms of weight k such that the q-expansion at one, and hence all the cusps of the curve $X_1(N)$, has coefficients in $\mathbb{Q}$. We denote the ring of integers of K by $\mathcal{O}_K$. Then we define $\mathcal{S}^k(\Gamma_1(N), \mathcal{O}_K)$, or more briefly $\mathcal{S}^k$ as both $\Gamma_1(N)$ and K will be fixed in our discussion, by:

$$\mathcal{S}^k := \bigoplus_{1 \leq j \leq k} S_j(\Gamma_1(N), K) \quad \cap \quad \mathcal{O}_K[[q]]$$

where of course we are identifying cusp forms with their q expansions at some fixed cusp. Then by definition we have, for any $k' \geq k$ inclusions $\mathcal{S}^k \hookrightarrow \mathcal{S}^{k'}$ and taking the union and p-adic completion we shall denote the resulting object by $\mathcal{S}$, i.e.,

$$\mathcal{S} := \overline{\bigcup_{k=1}^{\infty} \mathcal{S}^k}$$

where we take the completion with respect to the p-adic topology on $\mathcal{O}_K[[q]]$ given by the sup norm.

We as usual have the standard Hecke actions on all these spaces and we may consider the $\mathcal{O}_K$ algebra generated in the ring of endomorphisms of $\mathcal{S}^k$ by the Hecke operators T_n. We denote this algebra by $\mathcal{H}^k$. Then we may take the inverse limit of $\mathcal{H}^k$'s corresponding to the above inclusion maps. We denote the resulting object by $\mathcal{H}$. Now we need to indicate topologies on $\mathcal{S}$ and $\mathcal{H}$ to state a duality result between them (which may help in orienting the discussion a little). On $\mathcal{H}$ we put the natural inverse limit topology. On $\mathcal{S}$ we put the p-adic topology which is given by $||f|| = \sup_n a_n(f)$, where $f \in \mathcal{S}$ is $f = \sum_1^\infty a_n q^n$. Then the natural pairing:

$$\mathcal{H} \times \mathcal{S} \to \mathcal{O}_K$$

is given by $(T, f) = a_1(f|T)$. Then by Proposition III.1.2 of [G] we know that this pairing induces an isomorphism between $\mathcal{S}$ and the (continuous) dual of $\mathcal{H}$, i.e., $\mathrm{Hom}_{\mathcal{O}_K, conts}(\mathcal{H}, \mathcal{O}_K)$ in the notation of [G].

We have an embedding of the Iwasawa algebra $\mathbb{Z}_p[[X]]$ inside $\mathcal{H}$ which comes from the action of $\Gamma := 1 + p\mathbb{Z}_p$ on $\mathcal{S}$, cf. [H 3]. Denote a topological generator of Γ by γ. Then $\mathcal{H}/(\pi, \gamma - 1)$ naturally acts on $\widetilde{\mathcal{S}} := (\mathcal{S} \otimes \mathbb{F})^{\Gamma}$. Here $\mathbb{F}$ is the residue field of K and π is a uniformiser. The superscript Γ denotes taking fixed points under Γ. It follows from the description of the action of Γ in [H 3] that $\widetilde{\mathcal{S}}$ has as a quotient the space of mod p cusp forms of fixed level N and all weights (i.e., the subspace of $\mathbb{F}[[q]]$ given by the mod p reduction of $\mathcal{S}$). This space thus has the same definition as the space of modular forms mod p (for $\Gamma_1(N)$ and all weights) which we defined earlier in the section, with obvious modifications. In fact by a theorem of Katz (more precisely the variant of it for cuspidal forms) quoted as Theorem 1.1 in [H 3], it follows that $\widetilde{\mathcal{S}}$ coincides with the space of mod p cusp forms for $\Gamma_1(N)$. We may also note that the above pairing evidently induces the pairing we had defined earlier between mod p modular forms and the corresponding Hecke algebra.

The action of γ on $f \in S_k(\Gamma_1(N), K)$ is given by $f|\gamma = \gamma^k f$. This action extends continuously to $\mathcal{S}$. By the earlier paragraph $\mathcal{H}/(\pi, \gamma - 1)$ surjects onto the mod p Hecke algebra which will be the analog for the space of cusp forms of the H we defined earlier. Now pick any local component $\widetilde{R}$ of $\mathcal{H}$ such that the associated mod p representation is absolutely irreducible. The image of $\gamma - 1$ is not a zero divisor in the image of $\widetilde{R}$ in the ring of endomorphisms of $\mathcal{S} \otimes \mathbb{F}$ as we justify below. Note that $\widetilde{R}$ acts on this space through its quotient $\widetilde{R}/\pi\widetilde{R}$. But Proposition 2 (to be precise its natural (and identically proven) analog in the context of cusp forms) implies that the Krull dimension of a local component of the mod p Hecke algebra acting on the space of mod p cusp forms for $\Gamma_1(N)$ and of all weights is at least 2 assuming that the component is non-ordinary. Thus we see that we have proven the following theorem modulo the claim that the image of $\gamma - 1$ in the action of $\widetilde{R}$ on $\mathcal{S} \otimes \mathbb{F}$ is not a zero divisor.

Theorem 1. *If $\widetilde{R}$ is a non-ordinary local component of $\mathcal{H}$ such that the associated*

mod p representation is absolutely irreducible, then the Krull dimension of $\widetilde{R}$ is at least 4.

Proof. After what we have said we just need to prove that the image of $\gamma - 1$ in the quotient of $\widetilde{R}$ cut out by its action on $\mathcal{S} \otimes \mathbb{F}$ is not a zero divisor. We denote the space with which it is in topological duality with in $\mathcal{S}$ by $\mathcal{S}_R$. To prove our claim we need to check that if T in $\widetilde{R}$ is not 0 in its action on $\mathcal{S}_R \otimes \mathbb{F}$ then $T.(\gamma - 1)$ is also not 0. $\widetilde{R}/\pi\widetilde{R}$ acts on $\mathcal{S}_R \otimes \mathbb{F}$. There exists a non-zero element of this space which is not in the kernel of T. We may easily check that we may choose this non-zero element to be the image of some element f_k of $S_k(\Gamma_1(N), \mathcal{O}_K)$ in $\mathcal{S}_R \otimes \mathbb{F}$. Denote this image by $\tilde{f}_k$. Let n be the valuation of k with respect to p, i.e., $n = v_p(k)$. Then we choose k' such that it is congruent to k modulo $p-1$ and such that $v_p(k') \geq n+1$. Then we see easily that there is a $f_{k'}$ in $S_{k'}(\Gamma_1(N), \mathcal{O}_K)$ which is again in the part of the space in duality with $\widetilde{R}$ and congruent to f_k modulo p^{n+1}. We see this by multiplying by the Eisenstein series $E_{(p-1)p^n}$ (see [H 3]). Now it follows from [Ka] that f_k and $f_{k'}$ cannot be congruent modulo any higher power of p. We then consider the element $g := \frac{f_k - f_{k'}}{p^{n+1}}$. This is in $\mathcal{S}$. Then we see that the image of $g|T|(\gamma - 1)$ is not 0 in $\mathcal{S}_R \otimes \mathbb{F}$. For this we only need to note that γ acts on a form f of weight k by $f|\gamma = \gamma^k f$. As then we check that

$$g|T|\gamma - g|T = \frac{(\gamma^k - 1)f_k|T - (\gamma^{k'} - 1)f_{k'}|T}{p^{n+1}} \qquad (*)$$

is not 0 (in $\mathcal{S} \otimes \mathbb{F}$). As, by our choice of k', we see that $(*)$ is equal to a non-zero scalar multiple of $\tilde{f}_k|T$ in $\mathcal{S} \otimes \mathbb{F}$. We have thus shown that $\gamma - 1$ is not a zero divisor in the action of $\widetilde{R}$ on $\mathcal{S}_R \otimes \mathbb{F}$ and thus we are done.

We have given the alternative proof using weight variation that we had alluded to in §1.7 of Chapter 1.

2.3. Varying the level

Instead of varying the weight as above, we now fix the weight to be anything ≥ 2, say 2, but vary the p power level. In the same way as above, we can consider

the space of modular forms mod p of weight 2 for the 'group' $(\Gamma)' = \Gamma \cap \Gamma_0(p^\infty)$ identifying forms with their q-expansion, going mod p and so on. We assume that Γ has no p in its level. This time we have a filtration with respect to the p-power level the form comes from, i.e., we define $\ell(f) = r$ if the minimal p power level of a lift of f to a characteristic 0 form of weight 2 for the group Γ' is p^r. Just as before we have the operators U, V, Φ and though their meaning is different the effect on q-expansions is the same. We have that $ker(\Phi) = Im(V)$, $ker(U) = Im(\Phi)$, V is injective and U is surjective. Further we see that while U decreases the filtration of a form f if $\ell(f) > 1$, V always increases the filtration by 1. We see also that $\ell(f|\Phi) \leq \ell(f) + 1$ and in fact $\ell(f|\Phi) \leq l(f)$ if $\ell(f) \geq 2$. All this may be seen by using lemmas in the paper of Atkin-Lehner, cf. [A-L], and is in good analogy with what we saw in §2.2. We study non-ordinary local components of the Hecke algebra in this situation. So fix such a local component, say R. Just as before one has a decomposition of the kernel of U. Using this one can imitate the argument of Jochnowitz and prove that R is a power series ring in U over the subalgebra generated by the Hecke operators outside p. In this situation also we may try to find a Hecke operator T_n which is highly non semi-simple (in the sense of §2.2) in its action on the space of modular forms. We remark that in this setting the kernel of U grows systematically with level. One may easily see that for $r > 1$ we have that

$$\dim \ \ker(U|_{\widetilde{M_r}}) = \dim(\widetilde{M_r}) - \dim(\widetilde{M_{r-1}}).$$

Here by dim we mean dimension as $\overline{\mathbb{F}}_p$ vector space. As must be evident, we are using the natural notation that $\widetilde{M_r}$ denotes forms of filtration $\leq r$ in this context. In this case if we think about similar questions as in §2.2 and investigate the relationship between the two natural filtrations, i.e., one coming from the level filtration and the other from the degree of nilpotence with respect to the maximal ideal of a local component, we see that the local component being Noetherian immediately implies that the level filtration grows much faster upon using the dimension formula for the space of cusp forms of weight 2 and given level.

The fact analogous to §2.2 that the dimension of non-ordinary local components is at least 2 follows from Theorem 1 of the previous chapter immediately. Thus

Theorem 1 of that chapter may in fact be viewed as an analog of the theorem of Serre and Tate (which is for varying weights), quoted as lemma 3.4 in [J], in the context of varying levels. We may also prove this using Proposition 2 and the well known relation, see [H], between the Hecke algebras obtained by varying weight and varying p power level.

2.4. A comparison of filtrations on mod p forms

In studying local components of Hecke algebras acting on the space of mod p modular forms of fixed level and all weights, a question comes up which asks about relationships (if any) existing between two naturally occurring filtrations on the space of mod p forms. One filtration is the classical one due to Serre and Swinnerton-Dyer which comes from the least weight in which the q expansion of a mod p form occurs which we have discussed in §2.2. The other, which arises from looking at the mod p forms and writing them as the direct sum of spaces on which the maximal ideals of the Hecke algebra are respectively nilpotent, comes from the degree of nilpotence of the modular form lying in the space corresponding to the local component R, i.e., it is defined to be the least n such that m^n kills the form where m is the maximal ideal of the corresponding local component R (we shall always assume that R is non-Eisenstein in what follows). Then it is natural to ask what are the relations between these two filtrations. This question is of interest in determining algebraic properties of local components which have proven to be elusive. As far as the author knows, local components have been essentially studied in detail only in [J], though in the particular case of ordinary components, there is much more known even without reducing mod p due to the work of Hida, cf. [H]. There was little known about local components till Mazur introduced his idea of deformations. From his theory, as remarked before, it follows that the local components are Noetherian. It is somewhat surprising that this fact is not known without using deformations. The Noetherian property of local components has strong implications for the structure of mod p forms and can be used to study some exact sequences of modular forms which arise from the peculiar property of char p that there are inclusions between spaces of modular forms of different weights on

identifying forms with their q-expansion. We are in particular interested in studying the following exact sequence:

$$0 \to \widetilde{M_k} \to \widetilde{M_{k+p-1}} \to SS_k \to 0 \tag{1}$$

as a sequence of Hecke modules. It is a fact that we have noted before that this sequence is an essential extension of Hecke modules for k large enough. By this we just mean that $\widetilde{M_k}$ is maximal with the property that it contains $\widetilde{M_k}$ and the resulting exact sequence is split as a Hecke module. This is a direct consequence of the fact that there are only a finite number of eigensystems mod p. The study of this exact sequence we have come to believe holds the key to understanding the structure of local components of Hecke algebras. SS_k is defined by the exactness of the sequence. But there is another more direct interpretation of SS_k which depends on an idea of Serre which interpretes this, at least for k large enough, as functions on supersingular elliptic curves, cf. [E]. The mere existence of such a sequence as (1) produces a relationship between the filtrations we alluded to above. Thus it is an immediate consequence of the fact that SS_k has bounded dimension (which follows from the dimension formula for $S_k(\Gamma)$ for Γ any congruence subgroup of $SL_2(\mathbb{Z})$) that the part of $\widetilde{M_k}$ on which m is nilpotent is contained in $\widetilde{M[\mathrm{m}^{ak+b}]}$ for some constants a and b. We can make a and b explicit. It can for instance be easily checked that we may take a to be any number which bounds the dimensions of the SS_k's and we may take b to be the weight filtration of the eigenform corresponding to m, which is unique upto scalar. Henceforth by $\widetilde{M_k}$ we will mean those mod p forms of weight k on which m is locally nilpotent, for a fixed local component R and the corresponding maximal ideal m (so we are dropping the subscript R with which this space was adorned in §2.2). But to get information about local components we need also to be able to control the terms of the nilpotency filtration in terms of the weight filtration. It is unrealistic to expect that there is a similar linear relation in the other direction as it has been shown in §2.2 that the dimension of non-ordinary local components is at least 2 which makes this impossible. But we may modify our space $\widetilde{M_k}$ a little and look only at the part which is killed by the Atkin operator U. Then we may make the following conjecture.

Conjecture 1. There exist constants α and β such that we have the following inclusion for all n: $\mathcal{M}[\mathrm{m}^n] \hookrightarrow \mathcal{M}_{\alpha n+\beta}$.

Here by $\mathcal{M}$ as before we mean that part of the space which lies in the kernel of U as in §2.2 of this chapter. The reasons we have to make this conjecture are not very substantial but such as they are, arise from the perfect pairing between the local component R and the corresponding space of modular forms. If R is a non-ordinary local component, then U is transcendental in this component as noted in §2.2. We have a perfect pairing:

$$\mathcal{R} \times \mathcal{M} \to \overline{\mathbb{F}}_p.$$

Here $\mathcal{R}$ is by definition $R/(U)$. Then we expect that $\mathcal{R}$ has Krull dimension 1, motivated by the calculation of dimensions of deformation rings in certain cases (the unobstructed case) in [M] and some work of Hida [H 2]. We note that the above pairing induces a perfect pairing:

$$\mathcal{R}/\mathrm{m}^n \times \mathcal{M}[\mathrm{m}^n] \to \overline{\mathbb{F}}_p.$$

We would thus expect that the dimension of $\mathcal{M}[\mathrm{m}^n]$ grows linearly with n. This is the reason we have for making the above conjecture. To study this conjecture we return now to (1) and study some of its properties in greater detail. Thus with self-explanatory notation we have to study the exact sequence:

$$0 \to \mathcal{M}_k \to \mathcal{M}_{k+p-1} \to \mathcal{SS}_k \to 0 \tag{2}$$

as a sequence of $\mathcal{R}$ modules. We can make a related conjecture to the one stated above about this exact sequence.

Conjecture 2. There exists a constant x such that for any r the exact sequence:

$$0 \to \frac{\mathcal{M}_k}{\mathcal{M}_r} \to \frac{\mathcal{M}}{\mathcal{M}_r} \to \mathcal{C}_k \to 0 \tag{3}$$

is an essential extension of $\mathcal{R}$ modules for any $k \geq r + x$. Here $\mathcal{C}_k$ is defined by (3).

These two conjectures have an air of compatibility. At the moment both these conjectures are beyond us. We state and prove two propositions related to these conjectures.

Proposition 3. With notation as in Conjecture 2, (3) is an essential extension for k large enough (we are imagining that r is fixed).

Proof. We note that there exists a n such that $\mathcal{M}_r$ is contained in $\mathcal{M}[\mathrm{m}^n]$. Then the weight filtration being exhaustive, we note that as $\mathcal{R}$ is Noetherian, there exists a k such that $\mathcal{M}[\mathrm{m}^{n+1}]$ is contained in $\mathcal{M}_k$. Then we claim that this k works. As if (3) were not essential with this k, there exists a S strictly containing $\mathcal{M}_k$ such that the sequence corresponding to (3) made with S instead of $\mathcal{M}$ splits as a Hecke module. Then as $\mathcal{R}$ is commutative and as its maximal ideal m acts nilpotently on $\mathcal{M}$, we get that there is a non-zero f which is not in $\mathcal{M}_k$ but which is killed by m^{n+1}. This contradicts our choice of k.

We note that Proposition 3 is a very weak form of Conjecture 2.

Proposition 4. Conjecture 2 implies Conjecture 1.

Proof. We assume conjecture 2 and then prove the first conjecture by induction on the n of that conjecture. Let us assume that the eigenform corresponding to m has weight filtration y. Thus $\mathcal{M}[\mathrm{m}] \hookrightarrow \mathcal{M}_{\mathrm{y}}$. Now we assume that $\mathcal{M}[\mathrm{m}^n] \hookrightarrow \mathcal{M}_{xn+y}$. Then according to conjecture 2, the sequence:

$$0 \to \frac{\mathcal{M}_{x(n+1)+y}}{\mathcal{M}_{xn+y}} \to \frac{\mathcal{M}}{\mathcal{M}_{xn+y}} \to \mathcal{C}_{x(n+1)+y} \to 0$$

corresponds to an essential extension. But from this we may deduce that $\mathcal{M}[\mathrm{m}^{n+1}] \hookrightarrow \mathcal{M}_{x(n+1)+y}$ as if $f \in \mathcal{M}[\mathrm{m}^{n+1}]$, then the one-dimensional vector space spanned by the image of f in $\frac{\mathcal{M}}{\mathcal{M}_{xn+y}}$ is stable under the Hecke action and thus by what we have just said we conclude that conjecture 1 holds with the constants α and β being set equal to x and y respectively, completing thus the inductive step.

We are wont to believe that conjecture 2 (if it is true!) is less intractable than conjecture 1.

This proof of Proposition 3 will not give the explicit dependence of k on r which

we ask for in Conjecture 2. To make progress with these conjectures we essentially have to make quantitatively precise the fact that the action of $\mathcal{R}$ on $\mathcal{M}$ is highly non semi-simple. This gives a different approach than the one suggested in §2.2 to the problem of determining the dimension of local components of the mod p Hecke algebra.

We may even express a more outrageous hope. We note that $\mathcal{M}$ has the strong property that it has no submodules for $\mathcal{R}$ which are decomposable. This again follows from the fact that an eigenform for Hecke operators is determined by its eigenvalues upto scalar. We would conjecture that the same holds for $\mathcal{M}$ / $\mathcal{M}_k$ for large enough k. We may see easily that this implies both the conjectures above. As for proving this, we of course again have not a clue. We may also ask if the only infinite dimensional vector subspace of $\mathcal{M}$ which is stable under the Hecke action is the whole of $\mathcal{M}$. This would again imply that the Krull dimension of $\mathcal{R}$ is 1.

To conclude, in this chapter, besides the occasional affirmative results we prove, we have drawn a line between studying algebraic properties of local components of Hecke algebras and properties of sequences of Hecke modules. The lower bound we prove in Theorem 1 in this setting on the dimension of (non-ordinary) local components is very likely to be the right dimension. But we have not been able to prove that.

Chapter 3

On Fourier coefficients of eigenforms

3.1. Introduction.

It is a theorem of Deligne (and Deligne-Serre for weight 1) that for a cuspidal eigenform of the Hecke operators on the upper half plane which is of weight k, the eigenvalues of the Hecke operators T_p are algebraic integers a_p with $|a_p| \leq 2p^{(k-1)/2}$. In §3.2 of this chapter we pose a converse question to this, and analyse to what extent CM forms can be used to answer it. In §3.3 an analogous question is asked in the setting of Galois representations which can be thought of as the non-abelian analogue of the Grunwald-Wang theorem in Class Field Theory, and we answer it in one simple case. We may view these questions as asking for a kind of Chinese Remainder Theorem in the setting of automorphic and Galois representations respectively. The results of this chapter are obtained jointly with D. Prasad.

3.2. CRT for automorphic representations

The aim of this section is to pose the following question and provide an answer to it in some very particular cases.

Question 1: Suppose that we are given finitely many primes $p_1, \cdots, p_r$, and algebraic integers α_i for every $i, 1 \leq i \leq r$, which have the property that $\sigma(\alpha_i)\overline{\sigma(\alpha_i)} = {p_i}^{k-1}$ for some integer $k \geq 1$ and for every embedding $\sigma : \overline{\mathbb{Q}} \to \mathbb{C}$. Then does there exist a cusp form f of weight k which is an eigenform of all the Hecke operators such that the Euler factor at p_i of the L-series of f, for every i, $1 \leq i \leq r$, is

$$L_{p_i}(f,s) = \frac{1}{(1-\frac{\alpha_i}{{p_i}^s})(1-\frac{\overline{\alpha}_i}{{p_i}^s})}?$$

Assuming the Shimura-Taniyama-Weil conjecture according to which all ellip-

tic curves over $\mathbb{Q}$ are modular, this question can be settled rather easily in the affirmative in the case when $k = 2$ and $a_i = \alpha_i + \overline{\alpha}_i$ are rational integers as follows. By a theorem due to Honda and Tate, we can find an elliptic curves E_i over the finite fields $\mathbb{F}_{p_i}$ with p_i elements such that the cardinality of $E_i(\mathbb{F}_{p_i})$ is $1 + p_i - a_i$. If E is any elliptic curve whose reduction modulo p_i is the elliptic curve E_i for every i, $1 \leq i \leq r$, then the L-function of E is the Mellin transform of a desired modular form.

When $k = 2$ but a_i are not integers, we can't imitate the above proof even assuming the generalised form of the Shimura-Taniyama-Weil conjecture according to which abelian varieties with real multiplication over $\mathbb{Q}$ also arise as factors of the Jacobians of the modular curves $X_0(N)$. The problem being that it is not clear if we can lift an abelian variety with real multiplication over the finite field $\mathbb{F}_{p_i}$ to one over $\mathbb{Q}$. There is then the problem of doing this for finitely many primes $p_1, \cdots, p_r$ simultaneously. We, however, don't even know if an abelian variety over $\mathbb{F}_p$ can be lifted to one over $\mathbb{Q}$!

In this chapter we analyse to what extent CM forms can be used to answer the question. Here is the main result. All the numbers α_i appearing in the theorem below will have the property that $\sigma(\alpha_i)\overline{\sigma(\alpha_i)} = {p_i}^{k-1}$ for some integer $k \geq 2$ and for every embedding $\sigma : \overline{\mathbb{Q}} \to \mathbb{C}$.

Theorem 1. *Assume that $a_i = \alpha_i + \overline{\alpha}_i$ is an integer such that p_i does not divide a_i for any i, $1 \leq i \leq r$. Then there is a CM cuspidal eigenform f such that the Euler factor at p_i of the L-series of f is*

$$L_{p_i}(f, s) = \frac{1}{(1 - \frac{\alpha_i}{{p_i}^s})(1 - \frac{\overline{\alpha}_i}{{p_i}^s})}$$

if and only if the quadratic imaginary fields $K_i = \mathbb{Q}(\sqrt{a_i^2 - 4p_i^{k-1}})$ are independent of i.

Proof : We first recall that a CM modular form $f = f_\lambda$ is associated to a Größencharakter λ of a quadratic imaginary extension K of $\mathbb{Q}$. This Größencharakter λ can be thought of as a homomorphism $\lambda : I_K(c) \to \mathbb{C}^*$ (where $I_K(c)$ is the group

of fractional ideals prime to c where c is an ideal of K) such that for any $\alpha \epsilon \mathcal{O}_K$ with $\alpha \equiv 1 \pmod{c}$, where $\mathcal{O}_K$ is the ring of integers of K, $\lambda((\alpha)) = \alpha^a \overline{\alpha}^b$ for some integers a, b. As f_λ is a modular form, one moreover has $a \geq 0$, $b \geq 0$, and $ab = 0$. (One way, and not necessarily the best way, of seeing this last fact is to note that the Hodge-Tate type of representations attached to newforms is of the type $(*, *)$ where one of the $*$'s is zero. Then we note that the Galois representation attached to λ has this Hodge-Tate type if and only if $ab = 0$ as otherwise the representation will arise as a twist by a power of the norm character of a representation attached to a classical eigen cuspform and hence will not be Hodge-Tate of the required type (we refer to [Mi] for the fact that f_λ is an eigencuspform when $ab = 0$).)

The modular form f_λ is an eigenform of the Hecke operators and has the following Euler factor at primes p coprime to c:

$$L_p(f_\lambda, s) = \begin{cases} \frac{1}{(1-\lambda(\pi)p^{-s})} \frac{1}{(1-\lambda(\overline{\pi})p^{-s})}, & \text{if } (p) = \pi\overline{\pi} \\ \frac{1}{(1-\lambda(p)p^{-2s})}, & \text{if } (p) \text{ is inert} \\ \frac{1}{1-\lambda(\pi)p^{-s}}, & \text{if } (p) = \pi^2. \end{cases}$$

We now assume that the quadratic imaginary fields $K_i = \mathbb{Q}(\sqrt{a_i^2 - 4p_i^{k-1}})$ are all the same, say K, and in that case we construct a Größencharakter λ of K such that the associated modular form f_λ has the desired Euler factors at p_i, $1 \leq i \leq r$. We first note that as $k \geq 2$ and $p_i \not| a_i$, the prime ideal (p_i) splits in the quadratic imaginary field $K = K_i = \mathbb{Q}(\sqrt{a_i^2 - 4p_i^{k-1}})$ (as one can take the square root of $a_i^2 - 4p_i^{k-1}$ in $\mathbb{Q}_{p_i}$). Let $(p_i) = \pi_i \overline{\pi}_i$ be the factorisation of the ideal (p_i) in K as the product of prime ideals in K. Since $\alpha_i \overline{\alpha}_i = p_i^{k-1}$, and $\pi_i \overline{\pi}_i = (p_i)$, it follows from the assumption $p_i \not| a_i$ (possibly after replacing α_i by $\overline{\alpha}_i$) that $(\alpha_i) = \pi_i^{k-1}$, $(\overline{\alpha}_i) = \overline{\pi}_i^{k-1}$.

Let P_c denote the group of principal ideals (x) with $x \equiv 1 \pmod{c}$. Denote by μ_{00} the character on P_c given by $\mu_{00}((x)) = x^{k-1}$. (This is well defined for c large enough as the group of units of K is finite; moreover, c can be taken to be coprime to any given ideal which we take to be $\prod(p_i)$.) Let μ_0 be any extension of μ_{00} to $I(c)$. Our problem of the construction of λ will be solved as soon as we can demonstrate the existence of a Größencharakter λ which is unramified at π_i and

$\overline{\pi}_i$ for all i, $1 \leq i \leq r$, with $\lambda(\pi_i) = \alpha_i$, and $\lambda(\overline{\pi}_i) = \overline{\alpha}_i$ and whose infinity type is either $(a,0)$ or $(0,a)$ for some integer $a \geq 1$. From the relation $(\alpha_i) = \pi_i^{k-1}$, it follows that for the desired λ, $\lambda/\mu_0(\pi_i)$ and $\lambda/\mu_0(\overline{\pi}_i)$ must be roots of unity, say ω_i, ω_i'. Conversely if we can construct a Größencharakter ν which is unramified at π_i and $\overline{\pi}_i$ for all i, $1 \leq i \leq r$, with $\nu(\pi_i) = \omega_i$, and $\nu(\overline{\pi}_i) = \omega_i'$, then $\lambda = \nu\mu_0$ will be the desired Größencharakter. The existence of such a Größencharakter ν is a consequence of the theorem of Grunwald and Wang, cf. [A-T], completing this part of the theorem.

To prove that the fields K_i must be the same for the existence of a CM form f, it suffices to prove the following lemma.

Lemma 1. *Let f be a CM form such that the Euler factor at p of the L-series of f is $[1 - a_p p^{-s} + p^{k-1-2s}]^{-1}$. Assume that a_p is an integer with $p \nmid a_p$. Then f arises from a Größencharakter on the quadratic imaginary field $K = \mathbb{Q}(\sqrt{a_p^2 - 4p^{k-1}})$.*

Proof : Suppose that f arises from a Größencharakter λ on a quadratic imaginary field L. Looking at the Euler factor at p attached to the L-series of f, we find that p must split in L. Write the factorisation of (p) in L as $(p) = \pi\overline{\pi}$. Since the Euler factor at p of the L-series of f is $[1 - a_p p^{-s} + p^{k-1-2s}]^{-1}$, it follows that $\lambda(\pi) + \lambda(\overline{\pi}) = a_p$, and $\lambda(\pi)\lambda(\overline{\pi}) = p^{k-1}$. Therefore $\lambda(\pi)$ and $\lambda(\overline{\pi})$ lie in K. From the defining condition of a Größencharakter, it follows that there is an integer $h \geq 1$ such that $\lambda(\pi)^h \in L$. It can be checked that a power of $x + \sqrt{y}$ with x, y rational, $y \leq 0$, and $xy \neq 0$, is rational only if $x + \sqrt{y}$ is a rational multiple of the third root of unity w. It follows that $\lambda(\pi)^h$ is an element of K but not of $\mathbb{Q}$ if p does not divide a_p (we are using the condition $k \geq 2$ here). As $\lambda(\pi)^h$ lies in L, $K = L$.

The case when a_p is a non-zero integer but $p|a_p$ can't be obtained by CM forms as the next lemma shows. As the case when $a_p = 0$ can be obtained by any Größencharakter of any quadratic imaginary field in which (p) is inert, this completes all the cases in which CM forms can be used.

Lemma 2. *Let f be a CM form such that the Euler factor at p of the L-series of f is $[1 - a_p p^{-s} + p^{k-1-2s}]^{-1}$. Assume that a_p is a non-zero integer. Then p does not*

divide a_p.

Proof : Suppose that f arises from a Größencharakter λ on a quadratic imaginary field L. Looking at the Euler factor at p attached to the L-series of f, we find that p must split in L. Write the factorisation of (p) in L as $(p) = \pi\overline{\pi}$. Since the Euler factor at p of the L-series of f is $[1 - a_p p^{-s} + p^{k-1-2s}]^{-1}$, it follows that $\lambda(\pi) + \lambda(\overline{\pi}) = a_p$, and $\lambda(\pi)\lambda(\overline{\pi}) = p^{k-1}$. If $p|a_p$, then for all integers $h \geq 1$, $p|\lambda(\pi^h) + \lambda(\overline{\pi}^h)$.

Assume without loss of generality that the infinity type of λ is $(a, 0)$. Then there is an integer $h \geq 1$ such that $(\pi)^h$ is a principal ideal generated by, say γ, and such that

$$\lambda(\pi^h) = \gamma^a$$

and

$$\lambda(\overline{\pi}^h) = \overline{\gamma}^a.$$

Therefore $\gamma^a + \overline{\gamma}^a$ is divisible by p which is obviously not possible.

Remark 1 : The weight 1 case of Question 1 can be completely answered using CM forms. One simply has to take a quadratic imaginary field in which the prime ideals (p_i) split as $(p_i) = \pi_i\overline{\pi}_i$ and construct a finite order Größencharakter λ on L using the Grunwald-Wang theorem which is unramified at the primes π_i and $\overline{\pi}_i$, and has the property that $\lambda(\pi_i) = \alpha_i$, and $\lambda(\overline{\pi}_i) = \overline{\alpha}_i$ for every i, $1 \leq i \leq r$.

We also remark that one can ask a question related to Question 1 which has a negative answer. So we may fix a totally real algebraic integer, say α, and a positive integer N, and a prime p which does not divide N, and then ask if there exists a cuspidal eigenform, say f, of some weight $k > 1$, for the group $\Gamma_0(N)$, such that the eigenvalue of the p th Hecke operator T_p on f is α. Then the answer is no as the part of the Gouvea-Mazur conjectures already proven by Coleman [Co], implies that the "slopes" of the eigenvalues of the Atkin operator U_p, acting on the space

of cusp forms of all weights, for the group $\Gamma_0(Np)$, are discrete. Thus in particular there exists a number ε in the interval $(0,1)$, such that there are no "slopes" in the interval $(0,\varepsilon)$. Then any α with the property that its p-adic valuation, with respect to which the slopes have been measured, is in the interval $(0,\varepsilon)$, provides a negative answer to the question. We see this, as if there is a $f \epsilon S_k(\Gamma_0(N))$, $k > 1$, which is an eigenvector for T_p, with eigenvalue α, then at least one of the roots, which we will call a and b, of the equation $x^2 - \alpha x + p^{k-1}$, say a, has valuation in the interval $(0,\varepsilon)$. But then $f'(z) = f(z) - bf(pz)$, is an element of $S_k(\Gamma_0(Np))$, which is an eigenvector for U_p, with eigenvalue a. This contradicts the choice of ε. We refer to [Co] for the precise definition of "slopes" and more about the Gouvea-Mazur conjecture.

Remark 2 : There is by now a well-known result for automorphic representations, cf. Rogawski [Ro], that there are automorphic representations whose local components are pre-assigned discrete series representations at finitely many places. However, in question 1 we want to construct automorphic representations whose local components are pre-assigned unramified principal series at finitely many finite places, and a discrete series at infinity when $k \geq 2$. It is unlikely that this question can be handled by techniques of harmonic analysis alone, as it is essential to specify the data which is used to define the unramified principal series at the finitely many local places, in the situation of question 1, to be of arithmetic kind.

3.3. CRT for Galois representations

Here is the non-abelian version of the Grunwald-Wang theorem, and is the Galois theoretic analogue of question 1 for weight 1.

Question 2: Suppose that we are given semi-simple matrices $A_1, \cdots, A_r$ in $GL(n,\mathbb{C})$ such that the eigenvalues of A_i are roots of unity. Then is there a continuous irreducible representation $\Phi : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL(n,\mathbb{C})$ which is unramified at the primes p_i such that the conjugacy class of the image of the Frobenius at p_i under the representation Φ contains A_i for every i, $1 \leq i \leq r$?

Remarks :

3.1. If we do not insist on the irreducibility of the representation Φ, then such a representation can be easily constructed by the Grunwald-Wang theorem.

3.2. The answer to question 2 is no in the generality in which it has been posed here. The reason is that even though there are semi-simple matrices, say in $GL(2,\mathbb{C})$, for which the ratio of the eigenvalues are arbitrary large roots of unity, the finite subgroups of $GL(2,\mathbb{C})$ which act irreducibly on $\mathbb{C}^2$ are much more restricted.

3.3. One should therefore consider question 2 only for those matrices $A_1,\cdots,A_r$ which belong to a finite subgroup $G \subset GL(n,\mathbb{C})$ which acts irreducibly on $\mathbb{C}^n$. However, the example of Wang, cf. [A-T], shows that one may not be able to construct a representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ with values in G with the above local constraints.

3.4. We can ask more generally for the existence of a representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ with given restriction to the decomposition groups $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ which takes values in a finite subgroup $G \subset GL(n,\mathbb{C})$ for finitely many primes p.

At the moment we are unable to say anything about question 2, or its more general form in remark 3.4, except for the following proposition. In the following proposition, we have fixed embeddings of $\overline{\mathbb{Q}}$ in $\overline{\mathbb{Q}}_p$ for every prime p; we will abuse notation to include the prime at infinity also in the following proposition.

Proposition 1. *Let $G = S_n$, and suppose we are given $\rho_i : \mathrm{Gal}(\overline{\mathbb{Q}}_{p_i}/\mathbb{Q}_{p_i}) \to G$ for $1 \leq i \leq r$. Then there exists $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to G$ such that the restriction of ρ to $\mathrm{Gal}(\overline{\mathbb{Q}}_{p_i}/\mathbb{Q}_{p_i})$ is conjugate in G to ρ_i for every i.*

Proof : Let G_i denote the image in G of $\mathrm{Gal}(\overline{\mathbb{Q}}_{p_i}/\mathbb{Q}_{p_i})$ under ρ_i. Let X be the set $X = \{1,2,\cdots,n\}$ on which S_n, and therefore every G_i, operates. Write $X = \sqcup_\alpha X_{\alpha,i}$, a disjoint union, such that every $X_{\alpha,i}$ is invariant under G_i, and G_i operates transitively on the set $X_{\alpha,i}$. If $n_{\alpha,i}$ denotes the cardinality of $X_{\alpha,i}$, let $G_{\alpha,i}$ denote the image of G_i in the symmetric group $S_{n_{\alpha,i}}$. Therefore we have maps $\pi_{\alpha,i} : G_i \to G_{\alpha,i}$, and $\pi_i : G_i \to \prod_\alpha G_{\alpha,i}$.

Let K_i be the fixed field of the kernel of ρ_i so that K_i is a Galois extension of $\mathbb{Q}_{p_i}$

whose Galois group is canonically isomorphic to G_i. Let $K_{\alpha,i}$ denote the extension of $\mathbb{Q}_{p_i}$ contained in K_i which corresponds to the surjection $\pi_{\alpha,i} : G_i \to G_{\alpha,i}$. As $\pi_i : G_i \to \prod_\alpha G_{\alpha,i}$ is an injection, the compositum of $K_{\alpha,i}$ is K_i. Let $H_{\alpha,i} \subset G_{\alpha,i}$ denote the subgroup of $G_{\alpha,i}$ which is the stabiliser of an element (which will be arbitrarily chosen) of the set $X_{\alpha,i}$. Let $L_{\alpha,i}$ be the subfield of $K_{\alpha,i}$ fixed by $H_{\alpha,i}$. The degree of $L_{\alpha,i}$ over $\mathbb{Q}_{p_i}$ is $n_{\alpha,i}$. Let $f_{\alpha,i}$ denote an irreducible monic polynomial over $\mathbb{Q}_{p_i}$ of degree $n_{\alpha,i}$ one of whose roots generate $L_{\alpha,i}$. We assume, as we may, that the polynomials $f_{\alpha,i}$ are distinct for distinct α. Then $K_{\alpha,i}$ will be the splitting field of $f_{\alpha,i}$, and K_i will be the splitting field of the degree n polynomial $f_i = \prod_\alpha f_{\alpha,i}$ which has no multiple roots. Now let f be a polynomial over $\mathbb{Q}$ which approximates f_i well enough so that the roots of f generate the field extension K_i of $\mathbb{Q}_{p_i}$ and such that there is a matching of the roots of f with those of f_i over K_i such that the action of $\mathrm{Gal}(\overline{\mathbb{Q}}_{p_i}/\mathbb{Q}_{p_i})$ on the roots of f and f_i is the same after this identification. This is possible by an extension of Krasner's lemma which does this when f_i is irreducible. For the general case we claim that any monic polynomial f which is near enough to f_i also has factorisation $f = \prod f_\alpha$ with $\deg f_\alpha = \deg f_{\alpha,i}$, f_α irreducible monic and near to $f_{\alpha,i}$. For this it is enough to check that the mapping which takes the n-tuple consisting of the coefficients of f_α to the n-tuple consisting of the coefficients of f is an open mapping. Because of the open mapping theorem for $\mathbb{Q}_p^n$, it suffices to prove that the jacobian of such a mapping is non-zero at the point defined by $f_{\alpha,i}$. This is a simple consequence of the well-known fact that the mapping $(x_1, \cdots, x_n) \to (s_1, \cdots, s_n)$ where s_i is the i-th elementary symmetric function has non-zero jacobian at any point $(x_1, \cdots, x_n)$ with $x_l \neq x_k$ if $l \neq k$. This completes the proof of the claim from which we deduce that the roots of f_i and f generate the same field. Now using the roots of the degree n equation f, we get the desired map $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to S_n$ whose restriction to $\mathrm{Gal}(\overline{\mathbb{Q}}_{p_i}/\mathbb{Q}_{p_i})$ is conjugate in S_n to ρ_i for every i.

Remark 4: We don't know if the Proposition above is true even for $G = A_n$.

Remark 5: The problem of extending local representations to a global one is much

subtler than the problem of constructing extensions of global fields with given local extensions. This is evident even in the case of a global cyclic extension in which case when the local field extension is unramified extension of the same degree, the local representation will be the additional data specifying which generator of the cyclic group the Frobenius corresponds to.

References

[A-L] A. Atkin, J. Lehner, *Hecke operators on* $\Gamma_0(M)$, Math. Ann. 185 (1970), 134-160.

[A-Li] A. Atkin, W. Li, *Twists of newforms and pseudo-eigenvalues of* W*-operators*, Inv. Math. 48 (1978), 221-243.

[A-T] E. Artin, J. Tate, *Class Field Theory*, Benjamin, Reading, Mass. 1974.

[A-S] A. Ash, G. Stevens, *Modular forms in characteristic* ℓ *and special values of their L-functions*, Duke Math. J. 53 (1986), 849-868.

[B-L-R] N. Boston, H. W. Lenstra, K. Ribet, *Quotients of group rings arising from two-dimensional representations*, C. R. Math. 312 (1991), 323-328.

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

[C 1] H. Carayol, *Sur les représentations* ℓ*-adiques associées aux formes modulaires de Hilbert*, Ann. Sci. E. N. S. 19 (1986), 409-468.

[Co] R. Coleman, *p-adic Banach spaces*, preprint (1994).

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

[D-T] F. Diamond, R. Taylor, *Non optimal levels for mod* ℓ *modular representations of* $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$, Inv. Math. 115 (1994), 435-462.

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

[G] F. Gouvea, *Arithmetic of p-adic modular forms*, SLNM 1304.

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

[H] H. Hida, *Galois representations into $GL_2(\mathbb{Z}_p[[X]])$ attached to ordinary cusp forms*, Inv. Math. 85 (1986), 545-613.

[H 1] H. Hida, *Modular p-adic L functions and p-adic Hecke algebras*, to appear in Sugaku expositions.

[H 2] H. Hida, *Geometric modular forms*, CIMPA Summer School (1992).

[H 3] H. Hida, *Iwasawa modules attached to congruences of cusp forms*, Ann. Sci. E. N. S. 19 (1986), 231-273.

[J] N. Jochnowitz, *A study of the local components of the Hecke algebra mod ℓ*, Trans. AMS 270 (1982), 253-267.

[Ka] N. Katz, *Higher congruences between modular forms*, Ann. of Math. 101 (1975), 332-367.

[Ka 1] N. Katz, *A result on modular forms in characteristic p*, SLNM vol. 601, 53-61.

[L] S. Ling, *Congruences between cusp forms and geometry of Jacobians of modular curves*, Math. Ann. 295 (1993), 111-133.

[L-S] S. Ling, J. Oesterlé, *The Shimura subgroup of $J_0(N)$*, Astérisque 196-197 (1991), 171-203.

[M] B. Mazur, *Deforming Galois representations*, in Galois Groups over $\mathbb{Q}$, eds. Y. Ihara, K. Ribet, J. -P. Serre.

[M 1] B. Mazur, *Modular curves and the Eisenstein ideal*, Publ. Math., IHES 47 (1977), 33-186.

[M-R] B. Mazur, K. Ribet, *Two-dimensional representations in the arithmetic of modular curves*, Astérisque 196-197 (1991), 215-255.

[M-W] B. Mazur, A. Wiles, *Class fields of abelian extensions of* $\mathbb{Q}$, Inv. Math 76 (1984), 179-330.

[Ma] H. Matsumura, *Commutative ring theory*, CUP (1986).

[Mi] T. Miyake, *Modular forms*, Springer-Verlag, 1989.

[Mil] J. Milne, *Arithmetic Duality Theorems*, Perspectives in Mathematics, Academic Press (1986).

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

[R 1] K. Ribet, *On modular representations of* $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ *arising from modular forms*, Inv. Math. 100 (1990), 431-476.

[R 2] K. Ribet, *Report on mod* ℓ *representations of* $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$, in Motives, Part 2, ed. U. Jannsen et al., 639-676.

[R 3] K. Ribet, *Raising the levels of modular representations*, Prog. Math., vol. 81, Birkhauser (1990), 259-271.

[Ra] M. Raynaud, *Schémas en groupes de type* $(p, \cdots, p)$, Bull. Soc. Math. France, 102 (1974), 241-280.

[Ro] J. Rogawski, *Representations of GL(n) and division algebras over p-adic fields*, Duke Math. J. 50 (1983).

[S] G. Shimura, *Intoduction to the arithmetic theory of automorphic functions*, Princeton Univ. Press (1971).

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

[Se 1] J. -P. Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Inv. Math. 15 (1972), 259-331.

[Se 2] J. -P. Serre, *Trees*, Springer-Verlag (1980).

[Se 3] J. -P. Serre, *Le probleme des groupes de congruence pour* SL_2, Ann. of Math 92 (1970), 489-527.

[Se 4] J. -P. Serre, *Formes modulaires et fonctions zêta p-adiques*, SLNM vol. 350.

[T] R. Taylor, *On Galois representations associated to Hilbert modular forms*, Inv. Math. 98 (1989), 265-280.

[W] A. Wiles, *Modular elliptic curves and Fermat's Last Theorem*, preprint.