Invent. math. 124, 215–241 (1996)

Inventiones mathematicae

Classical and overconvergent modular forms

Robert F. Coleman

Department of Mathematics, University of California at Berkeley, Berkeley, CA 94720, USA

Oblatum 20-VII-1993 & 27-X-1994

to Reinhold Remmert

The purpose of this article is to use rigid analysis to clarify the relation between classical modular forms and Katz's overconvergent forms. In particular, we prove a conjecture of Gouvêa [G, Conj. 3] which asserts that every overconvergent p-adic modular form of sufficiently small slope is classical. More precisely, let $p > 3$ be a prime, K a complete subfield of $\mathbf{C}_p$, N a positive integer such that $(N, p) = 1$, and k an integer. Katz [K-pMF] has defined the space $M_k(\Gamma_1(N))$ of overconvergent p-adic modular forms of level $\Gamma_1(N)$ and weight k over K (see Sect. 2) and there is a natural map from weight k modular forms of level $\Gamma_1(Np)$ with trivial character at p to $M_k(\Gamma_1(N))$. We will call these modular forms classical modular forms. In addition, there is an operator U on these forms (see [G-ApM, Chap. II, Sect. 3]) such that if F is an overconvergent modular form with q-expansion $F(q) = \sum_{n \geqq 0} a_n q^n$ then

$$UF(q) = \sum_{n \geqq 0} a_{pn} q^n \ .$$

(In fact, all this exists even when $p = 2$ or 3 (see [K-pMF] or [C-BFM]).) We prove, Theorem 6.1, that if F is a generalized eigenvector for U with eigenvalue λ (i.e., in the kernel of $(U - \lambda)^n$ for some positive integer n) of weight k and λ has p-adic valuation strictly less than $k - 1$, then F is a classical modular form. In this case the valuation of λ is called the slope of F. In the case when F has slope 0, this is a theorem of Hida [H] and, more generally, it implies Gouvêa's conjecture mentioned above (which is the above conclusion under the additional hypothesis that the slope of F is at most $(k-2)/2$). This almost settles the question of which overconvergent eigenforms are classical, as the slope of any classical modular form of weight k is at most $k - 1$. In Sect. 7, we investigate the boundary case of overconvergent modular forms of slope one less than the weight. We show that non-classical forms with this property exist but that any eigenform for the full Hecke algebra of weight $k > 1$ is classical if it does not equal $\theta^{k-1}G$ (see below) where

G is an overconvergent modular form of weight $2-k$. In Sect. 8, we prove a generalization of Theorem 6.1, Theorem 8.1, which relates forms of level $\Gamma_1(Np)$ to what we call overconvergent forms of level $\Gamma_1(Np)$ and in Sect. 9, we interpret these latter as certain Serre p-adic modular forms with non-integral weight [S].

The central idea in this paper is expressed in Theorem 5.4 which relates overconvergent modular forms to the de Rham cohomology of a coherent sheaf with connection on an algebraic curve. More precisely, we show that there is a map θ^{k+1} for non-negative k from modular forms of weight $-k$ to modular forms of weight $k+2$ which on q-expansions is $(qd/dq)^{k+1}$. When $N>4$, the kth symmetric power of the first relative de Rham cohomology of the universal elliptic curve with a point of order N over the modular curve $X_1(N)$ is naturally a sheaf with connection. Theorem 5.4 is the assertion that the cokernel of θ^{k+1} is the first de Rham cohomology group of restriction of this sheaf to the complement of the zeros of the modular form E_{p-1} on $X_1(N)$.

The above result, Theorem 6.1, is intimately connected with the conjectures of Gouvêa and Mazur on families of modular forms in [GM-F] and [GM-CP]. Indeed, in a future article [C-BFM] we will use it to deduce qualitative versions of these conjectures. (We will also explain how to handle $p=2$ or 3 in [C-BFM].)

1. The rigid subspaces associated to sections of invertible sheaves

Let v denote the complete valuation on the p-adic numbers $\mathbf{Q}_p$ such that $v(p)=1$ and let $\mathbf{C}_p$ denote the completion of an algebraic closure of $\mathbf{Q}_p$ with respect to the extended valuation (which we still call v). We also fix a non-trivial absolute value $|\ |$ on $\mathbf{C}_p$ compatible with v. Suppose R is the ring of integers in a complete discretely valued subfield K of $\mathbf{C}_p$.

Suppose $\mathscr{X}$ is a reduced proper flat scheme of finite type over R and $\mathscr{L}$ is an invertible sheaf on $\mathscr{X}$. Let s be a global section of $\mathscr{L}$. Suppose x is a closed point of the subscheme $X := \mathscr{X}\otimes K$ of $\mathscr{X}$. Let K_x denote the residue field of x, which is a finite extension of K, so the absolute value on K extends uniquely to K_x, and let R_x denote the ring of integers in K_x. Then, since $\mathscr{X}$ is proper, the morphism $Spec(K_x)\to X$ corresponding to x extends to a morphism $f_x : Spec\ (R_x)\to\mathscr{X}$. Since K is discretely valued, $f_x^*\mathscr{L}$ is generated by a section t. Let $f_x^*s = at$ where $a\in R_x$ we set $|s(x)|=|a|$. This is independent of the choice of t. We will, henceforth regard X as a rigid space over K. We claim, for each $r>0$ such that $r\in|\mathbf{C}_p|$, there is a unique rigid subspace X_r which is a finite union of affinoids whose closed points are the closed points x of X such that $|s(x)|\geqq r$. Indeed, there exists a finite affine open cover $\mathscr{C}$ of $\mathscr{X}$ such that the restriction of $\mathscr{L}$ to Z for each Z in $\mathscr{C}$ is trivial. For each $Z\in\mathscr{C}$, let t_Z be a generator of $\mathscr{L}(Z)$ and suppose $s|_Z = f_Z t_Z$ where $f_Z\in\mathscr{O}_X(Z)$. Let $\hat{Z}$ denote the fiber product of the formal completion of Z along its special fiber and $Spec(K)$. Then $\hat{Z}$ is an affinoid over K and $X_r\cap\hat{Z}=\{x\in\hat{Z} : |f_Z(x)|\geqq r\}$. Since this is known to be the set of points of an affinoid [BGR, Sect. 7.2], we have established our claim. Now, for $r \in \mathbf{R}$, $r \geqq 0$, the set of closed points of X such that $|s(x)| > r$ is also the set of closed points of a rigid space $X_{(r)}$ as $\{X_u : u \in |\mathbf{C}_p|, u > r\}$ is an admissible cover.

Alternatively, if L is the line bundle whose sheaf of sections is $\mathscr{L}$ then there is a natural metric on L and if we regard s as a section of $L \to X$, X_r is the pullback of the rigid subspace consisting of points whose absolute value is greater than or equal to r. Or, if $a \in R$ and $|a| = r$, then X_r may be identified with the fiber product over R of $Spec(K)$ and the p-adic completion of $Spec_X(Sym(\mathscr{L})/(s-a))$. (One can deal with sections of locally free sheaves just as well.)

Now suppose X is an irreducible curve and $r \in |\mathbf{C}_p|$. Then, either X_r is an affinoid or $X_r = X$, because such is true for any finite union of affinoids in an irreducible curve. If $\mathscr{L} \neq s\mathscr{O}_X$, the reduction $\bar{s}$ of s is not zero and $r \neq 0$, then X_r is an affinoid. In fact, if X is smooth, $\bar{s}$ has only zeros of multiplicity one and $1 \geqq r > 0$ we claim X_r is the complement of a finite union of wide open disks. Indeed, suppose all the zeros of s are defined over R (this is not really necessary, it just makes things easier to visualize). Let $\mathscr{C}$ etc. be as above and let $Z \in \mathscr{C}$. Then t_Z is a local parameter at Q for each zero Q of s in Z and, in particular, the restriction t_Q of t_Z to the residue disk containing Q gives an isomorphism onto the unit disk $B(0, 1)$. Moreover, $X_r \cap Z = Z - \bigcup t_Q^{-1} B(0, r)$. This establishes the above claim. It follows that $X_{(r)}$ is the complement of a finite union of affinoid disks and so is a wide open by definition. (See [RLC] and also [Sch].) Such spaces are quasi-Stein spaces [Ki].

2. Application to overconvergent modular forms

Let $p > 3$ be a prime and let $N > 4$ be an integer such that $(p, N) = 1$. Let X denote the model with good reduction of $X_1(N)$ over R and C the subscheme of cusps. (We will also use C to denote the degree of the divisor C when no confusion will arise.) Let $E \to X$ denote the universal generalized elliptic curve with $\Gamma_1(N)$ structure, $f : E^{\mathrm{sm}} \to X$ the subscheme of E consisting of points smooth over X and $\tilde{C} = f^{-1}C$. Let $\Omega^1_{E^{\mathrm{sm}}/X}(\log \tilde{C}) = \Omega^1_{E^{\mathrm{sm}}}(\log \tilde{C})/f^*\Omega^1_X(\log C)$ and $\omega = f_*\Omega^1_{E^{\mathrm{sm}}/X}(\log \tilde{C})$. Now ω is an invertible sheaf and we have a section E_{p-1} of $\omega^{\otimes p-1}$. Since the reduction of E_{p-1} vanishes simply at each supersingular point, the rigid spaces $X_{(r)} = \{x : |E_{p-1}(x)| > r\}$ are each the complement of a union of closed disks, one in each supersingular residue disk, by the discussion of the previous section.

Let $Z = X_1$, $W_1 = X_{(p^{-p/(p+1)})}$ and $W_2 = X_{(p^{-1/(p+1)})}$. Then $W_2 \subset W_1$ and Z, the ordinary locus, is the unique minimal underlying affinoid (see [RLC]) of either W_1 or W_2 containing the cusps. Let $M_k =: M_k(\Gamma_1(N)) =: \omega^k(W_1)$ for $k \in \mathbf{Z}$. Then M_k may be described in terms of Katz's overconvergent forms of weight k. Indeed, if $r \in R$, $\omega^k(X_{|r|})$ may be identified with $S(R, r, N, k) \otimes K$ (see [K-pMF, Sect. 2.9]) and $M_k = \varprojlim_s \omega^k(X_s)$ where s approaches $p^{-p/(p+1)}$

from above. We call the sections of ω^k on Z *convergent* modular forms and those of X_s for any $s < 1$ *overconvergent* modular forms.

Remark. We point out that Katz's overconvergent modular forms are only defined for integral weights while Serre's p-adic forms may have weights in $\mathbf{Z}_p \times \mathbf{Z}/(p-1)\mathbf{Z}$. Katz discusses the relationship between the two types of objects in [K-pMF, Sect. 4.5]. In particular, he shows that Serre's forms of weight (k, k) for integral k are his convergent forms of weight k. In a future article, we will introduce a notion of "overconvergent " p-adic modular forms with weight in $B(0, |\pi/p|) \times \mathbf{Z}/(p-1)\mathbf{Z}$ which incorporates the forms of both Katz and Serre (see also Sect. 9).

Let E_i denote the pullback of E^{sm} to W_i. Katz describes [K-pMF, Sect. 3.10] a commutative diagram

$$\begin{array}{ccc} E_2 & \xrightarrow{\Phi} & E_1 \\ \downarrow & & \downarrow \\ W_2 & \xrightarrow{\phi} & W_1 \end{array} \tag{2.1}$$

where Φ and ϕ are finite morphisms. There is a canonical family of subgroup schemes $\mathscr{K}$ of rank p over W_1 in the family of elliptic curves E_1 over W_1 and ϕ is a morphism such that the family $E_2/\mathscr{K}_{E_2}$ over W_2 is canonically isomorphic to the pullback $E_1^{(\phi^*)}$ of E_1 to W_2. Then Φ is the composition of the isogeny $\pi : E_2 \to E_2/\mathscr{K}_{E_2}$ and the natural projection from $E_1^{(\phi^*)}$ to E_1.

Using this we will make ϕ^*-linear transformations $V_k : \mathscr{H}_k(W_2) \to \mathscr{H}_k(W_1)$ where $\mathscr{H}_k = Sym^k \mathbf{R}^1 f_* \Omega^{\boldsymbol{\cdot}}_{E/X}(\log \tilde{C})$ for $k \in \mathbf{Z}$, $k \geqq 0$ (note that $\mathscr{H}_0 = \mathscr{O}_X$). We first observe that the complex $\Omega^{\boldsymbol{\cdot}}_{E_2/W_2}(\log \tilde{C})$ is naturally isomorphic to the complex $\Phi^* \Omega^{\boldsymbol{\cdot}}_{E_1/W_1}(\log \tilde{C})$. What this means is that we get a natural map of complexes

$$\Phi_* \Omega^{\boldsymbol{\cdot}}_{E_2/W_2}(\log \tilde{C}) \to \Omega^{\boldsymbol{\cdot}}_{E_1/W_1}(\log \tilde{C})$$

which takes $g\Phi^*(\alpha)$ to $Trace_{\Phi}(g)\alpha$ where g is a section of $\mathscr{O}_{E_2}(\Phi^{-1}V)$ and α is a section of $\Omega^1_{E_1/W_1}(V)$ for an open set V of E_1. Second, as

$$Sym^k \mathbf{R}^1 f_* \Phi_* \Omega^{\boldsymbol{\cdot}}_{E_2/W_2}(\log \tilde{C}) \cong \phi_* Sym^k \mathbf{R}^1 f_* \Omega^{\boldsymbol{\cdot}}_{E_2/W_2}(\log \tilde{C}) ,$$

we get a map

$$V_k : \mathscr{H}_k(W_2) \to \mathscr{H}_k(W_1) .$$

Similarly we get a map

$$(\Omega^1_{X_1(N)} \otimes \mathscr{H}_k)(W_2) \to (\Omega^1_{X_1(N)} \otimes \mathscr{H}_k)(W_1)$$

which we also call V_k.

We have a natural map of complexes $\Phi^* : \Phi^{-1}\Omega^{\boldsymbol{\cdot}}_{E_1/W_1} \to \Omega^{\boldsymbol{\cdot}}_{E_2/W_2}$. This yields maps $\mathscr{H}_k(W_1) \to \mathscr{H}_k(W_2)$ and

$$(\mathscr{H}_k \otimes \Omega^1_{X_1(N)})(W_1) \to (\mathscr{H}_k \otimes \Omega^1_{X_1(N)})(W_2)$$

which we denote by F_k.

Now we have the Gauss–Manin connection ∇_k with log-poles at C on $\mathscr{H}_k$. That is, $\nabla_k : \mathscr{H}_k \to \Omega^1_X(\log C) \otimes \mathscr{H}_k$ and it is easy to see that

$$\nabla_k \circ V_k = V_k \circ \nabla_k \quad \text{and} \quad \nabla_k \circ F_k = F_k \circ \nabla_k \tag{2.2}$$

as maps from $\mathscr{H}_k(W_2)$ to $(\Omega^1_X(\log C) \otimes \mathscr{H}_k)(W_1)$ and from $\mathscr{H}_k(W_1)$ to $(\Omega^1_X \otimes \mathscr{H}_k)(W_2)$ respectively.

We can also describe V_k as the composition

$$\mathscr{H}_k(W_2) \stackrel{(\check{\pi}^*)^k}{\to} Sym^k \mathbf{R}^1 f_* \Omega^{\bullet}_{E_1^{(\phi^*)}/W_2}(W_2) = (\phi^* \mathscr{H}_k)(W_2) \stackrel{Tr_\phi}{\to} \mathscr{H}_k(W_1) \tag{2.3}$$

and F_k as the composition

$$\mathscr{H}_k(W_1) \to (\phi^* \mathscr{H}_k)(W_2) = Sym^k \mathbf{R}^1 f_* \Omega^{\bullet}_{E_1^{(\phi^*)}/W_2}(W_2) \stackrel{(\pi^*)^k}{\to} \mathscr{H}_k(W_2)\,,$$

$$h \mapsto h^{(\phi^*)}\,.$$

From this it is easy to see that if $h \in \mathscr{H}_j(W_2)$, $g \in \mathscr{H}_k(W_1)$ then

$$V_{k+j}(hF_k(g)) = p^k V_j(h) g\,. \tag{2.4}$$

This is also true with $\mathscr{H}$ replaced by $\Omega^1_{X_1(N)} \otimes \mathscr{H}$.

Remarks. The restriction of $\mathscr{H}_k$ to Z (or better the dagger completion of Z) is an F-crystal in the sense of Katz [K-TD] if one takes the map from $\phi^* \mathscr{H}_k$ to $\mathscr{H}_k$ to be $(\pi^*)^k$.

We will, henceforth, use $\check{\pi}^*$ to denote $\left(\check{\pi}^*\right)^k$.

3. Hodge and U

If $k \geqq 0$, the Hodge filtration on $\mathscr{H}_k$ is a descending filtration

$$Fil^0 \mathscr{H}_k = \mathscr{H}_k \supseteq \ldots \supseteq Fil^{k+1} \mathscr{H}_k = 0$$

such that for $0 \leqq i \leqq r$ and $0 \leqq j \leqq s$

$$Fil^i \mathscr{H}_r \cdot Fil^j \mathscr{H}_s = Fil^{i+j} \mathscr{H}_{r+s} \tag{3.1}$$

as coherent sheaves. It is clear that V_k and F_k respect these filtrations.

We observe that for $k \geqq 1$, $\mathscr{H}_k$ is self-dual with respect to a natural inner product $\langle , \rangle_k : \mathscr{H}_k \times \mathscr{H}_k \to \mathscr{O}_{X_1(N)}$ which when $k = 1$, away from the cusps, is just the cup product and more generally satisfies

$$\langle f_1 f_2 \ldots f_k, g_1 g_2 \ldots g_k \rangle_k = \prod_{1 \leqq i,j \leqq k} \langle f_i, g_j \rangle_1$$

for f_i, g_i local sections of $\mathscr{H}_1$. This leads, in particular, to the exact sequence

$$0 \to \omega \to \mathscr{H}_1 \to \omega^{-1} \to 0\,, \tag{3.2}$$

which together with (3.1) implies $Gr_i\mathscr{H}_k = Fil^{k-i}\mathscr{H}_k/Fil^{k+1-i}\mathscr{H}_k$ is canonically isomorphic to ω^{k-2i}.

We will identify M_j with $Gr_0\mathscr{H}_j(W_1)$ when $j \geqq 0$ and with $Gr_{-j}\mathscr{H}_{-j}(W_1)$ when $j < 0$. We let $U_{(j)}$ be the operator on M_j

$$V_j \circ \frac{1}{p} Res^{W_1}_{W_2} \text{ when } j \geqq 0 \quad \text{ and } \; p^j V_{-j} \circ \frac{1}{p} Res^{W_1}_{W_2} \text{ when } j < 0\,.$$

We will frequently drop the subscript (j) from $U_{(j)}$ when the context makes it clear on which space we are acting and sometimes abuse notation and allow $U_{(j)}$ to mean $(1/p)V_j|_{\omega^j(W_2)}$ when $j \geqq 0$.

Remark. The operator $U_{(j)}$ extends to any of the spaces $\omega^j(X_s)$ for any $1 \geqq s > p^{p/(p+1)}$ but any overconvergent eigenvector of $U_{(k)}$ analytically continues to W_1. (See the proof of [G, Cor. II.3.18].) The value of considering these larger spaces is that for $s \in |K|$ they are Banach spaces and one can apply the theory of Serre [S-B].

Suppose $g \in M_k$. We set $\sigma(g)$ equal to $F_k(g)/p^k = \pi^* g^{(\phi^*)}/p^k$ when $k \geqq 0$ and $F_k(g)$ when $k < 0$. So that $\sigma(g) \in \omega^k(W_2)$. It then follows from (2.4) that if $h \in \omega^j(W_2)$ and $kj \geqq 0$,

$$U_{(k+j)}(h\sigma(g)) = U_{(j)}(h)g\,. \tag{3.3}$$

In particular, if $h \in M_k$ and a is a rigid function on W_2 then

$$U(a\sigma(h)) = \frac{1}{p}\, Trace_{\phi}(a)h\,. \tag{3.4}$$

Equation (3.3) will follow, in general, from the following proposition.

Remark. The map σ is what is called φ in [K-pMF, Sect. 3] (and *Frob* in [G, Chap. 2, Sect. 2]), which is only defined there when $k \geqq 2$ in general and when $k = 1$ in some cases. Indeed, we may regard an element h of M_k as a function which assigns to pairs $(G/B,\omega)$ where B is a K algebra, G is a fiber of E_1/W_1 over a B valued point of W_1 and ω is a differential on G which generates the invariant differentials on G over B, an element $h(G/B,\omega)$ of B by the rule $h|_G = h(G/B,\omega)\omega^k$. Then if G/B is the fiber over a B valued point s of W_2,

$$\begin{aligned} F_k(h)|_G &= \pi^*(h((G/\mathscr{K}_G)/B, \check{\pi}^*\omega)(\check{\pi}^*\omega)^k) \\ &= p^k h((G/\mathscr{K}_G)/B, \check{\pi}^*\omega)\omega^k\,, \end{aligned}$$

where we regard $G/\mathscr{K}_G$ as the fiber over the B valued point $\phi \circ s$ of W.

Proposition 3.1 *Suppose $k \geqq 0$. The map induced by $V_k \circ Res^{W_1}_{W_2}$ on $Gr_i\mathscr{H}_k(W_1) \cong M_{k-2i}$ is $p^{i+1}U$ for $0 \leqq i \leqq k$.*

Proof. We pass to the underlying affinoid Z and repeat all previous constructions in this context. This gives us the advantage of not having to worry about the fact that $W_1 \nsupseteq W_2$. We will let A denote the pullback of E to Z.

We will now follow Appendix 2 of [K-pMF]. (Note: What is denoted by the symbol φ there is what is called ϕ^* here.)

Suppose v is an invariant differential on A generating ω. Then $\check{\pi}^* v = \lambda v^{(\phi^*)}$ and $\check{\pi}^* v^{-1} = (p/\lambda)(v^{-1})^{(\phi^*)}$ for some invertible λ. Suppose f is a section of $Fil^{k-i}\mathscr{H}_k$ such that $f \equiv a\sigma(h)(A,v)v^{k-i} \cdot v^{-i}$ modulo $Fil^{k+1-i}\mathscr{H}_k$ where h is a weight $k-2i$ modular form and a is a rigid function on Z. (Here we are using (3.1).) Then,

$$\begin{aligned}\check{\pi}^*(f) &\equiv a\sigma(h)(A,v)(\check{\pi}^* v)^{k-i} \cdot \check{\pi}^*(v^{-i})\\ &\equiv a\sigma(h)(A,v)p^i\lambda^{k-2i}(v^{k-i})^{(\phi^*)} \cdot (v^{-i})^{(\phi^*)} \bmod Fil^{k+1-i}\mathscr{H}_k\,.\end{aligned}$$

Now then,

$$V_k(f) \equiv p^i\, Trace_\phi(\lambda^{k-2i}a\sigma(h)(A,v))v^{k-i} \cdot v^{-i} \bmod Fil^{k+1-i}\mathscr{H}_k\,,$$

using (2.3), while

$$\begin{aligned}\sigma(h)(A,v) &= h(A/\mathscr{K}_Z, \check{\pi}^* v)\\ &= h(A^{(\phi^*)}, \lambda v^{(\phi^*)})\\ &= \lambda^{2i-k}\phi^*(h(A,v))\,.\end{aligned}$$

Thus

$$V_k(f) \equiv p^i\ Trace_\phi(a)h(A,v)v^{k-i} \cdot v^{-i} \bmod Fil^{k+1-i}\mathscr{H}_k$$

and so, on Gr_i, V_k acts as $p^{i+1}U$ using (3.4). □

4. Kodaira–Spencer and the theta operator

We have a Kodaira–Spencer map $\kappa : \omega^2 \to \Omega^1(\log C)$ of coherent sheaves on set $\nabla = \nabla_1$. $X_1(N)$ defined as follows: If β and v are two local sections of ω we set

$$\kappa(\beta \otimes v) = \langle \beta, \nabla v\rangle\,,$$

where on the right we regard β and v as sections of $\mathscr{H}_1$. This map is an isomorphism. More generally, we have an injection of sheaves $\kappa_k : \omega^{k+2} \to \Omega^1(\log(C)) \otimes \mathscr{H}_k$ determined by the correspondence

$$\kappa_k\ :\ \beta \otimes v \otimes \eta \mapsto \kappa(\beta \otimes v) \otimes \eta\,,$$

where η is a local section of ω^k and on the right we regard it as a local section of $\mathscr{H}_k$.

Proposition 4.1

$$V_k \circ \kappa_k = \kappa_k \circ U_{(k+2)}\,.$$

Proof. Suppose $\eta \in \omega^k(W_2)$, $\beta, v \in M_1$ and $f = \sigma(\beta \otimes v) \otimes \eta$. Then on the one hand,

$$U_{(k+2)}(f) = (\beta \otimes v) \otimes U_k(\eta)\,.$$

On the other hand,

$$\begin{aligned} V_k(\kappa_k f) &= V_k(\langle \sigma\beta, \nabla\sigma v\rangle \otimes \eta) \\ &= V_k(\langle \pi^*\beta^{\phi^*}/p, \nabla\pi^* v^{\phi^*}/p\rangle \otimes \eta) \\ &= V_k(\langle \pi^*\beta^{\phi^*}/p, \pi^*\nabla^{\phi^*} v^{\phi^*}/p\rangle \otimes \eta) \\ &= (1/p)V_k(\phi^*\langle \beta, \nabla v\rangle \otimes \eta) = (1/p)\langle \beta, \nabla v\rangle \otimes V_k(\eta)\,. \end{aligned}$$

The proposition follows from this and the definitions. □

Lemma 4.2 *The map*

$$Fil^1\mathscr{H}_k \xrightarrow{\nabla} (\Omega^1_{X_1}(N)(\log C)\otimes \mathscr{H}_k)/(\Omega^1_{X_{1(N)}}(\log C)\otimes Fil^k\mathscr{H}_k)$$

is an isomorphism of sheaves of vector spaces.

Proof. The map $\mathscr{H}_k \xrightarrow{\nabla} \Omega^1_{X_1}(N)(\log C)\otimes\mathscr{H}_k$ shifts the filtrations by 1 (Griffiths transversality) and induces isomorphisms on the graded pieces by what we know about the Kodaira–Spencer map. □

This implies that we get a natural map $M_{-k} \to M_{k+2}$. Indeed, let $w \in M_{-k}$. Lift it to a section $\tilde{w}$ of $\mathscr{H}_1$. Let s be a section of $Fil^1\mathscr{H}_k$ on W_1 such that $\nabla\tilde{w} \equiv \nabla s$ modulo $(\Omega^1_{X_{1(N)}}(\log C)\otimes(Fil^k\mathscr{H}_k))(W_1) = M_{k+2}$. Then the map is $w \mapsto \nabla(\tilde{w}-s)$. At the cusp ∞, one computes (see [pSI, Sect. 9]) that this map is

$$f(q) \mapsto (-1)^k\theta^{k+1}f(q)/k!\,,$$

where $\theta = qd/dq$. In particular,

Proposition 4.3 *There is a linear map from M_{-k} to M_{k+2} which on q-expansions is θ^{k+1}.*

We will, henceforth, denote this map by the expression θ^{k+1}.

Katz tells us [K-pMF, Appendix 1] that we can “canonically, but not functorially”, regard $\mathscr{H}_k$ for $k \geqq 0$, as

$$\omega^k \oplus \omega^{k+2} \oplus \ldots \oplus \omega^{-k}\,. \tag{4.1}$$

Suppose $f \in M_k$. Suppose first $k \geqq 0$. By the above we can consider it as a section of $\mathscr{H}_k$. Then ∇f is a section of $\mathscr{H}_k \otimes \Omega^1_X(\log C)$. By virtue of Katz’s decomposition (4.1), we can project onto $(\omega^k \otimes \Omega^1_X(\log C))(W_1)$ which by virtue of Kodaira–Spencer isomorphism we can identify with M_{k+2}. Call this element $\delta_k f$. Now suppose $k < 0$. By (4.1), we can regard f as an element of $\mathscr{H}_{-k}(W_1)$. Then the projection of $\nabla_{-k} f$ onto $M_k \otimes \Omega^1_X(\log C)(W_1)$ can be regarded as an element of M_{k+2} and we call this element $\delta_k f$. In either case, calculating at ∞ we find

$$(\delta_k f)(q) = \theta(f(q)) - kf(q)E_2(q)/12\,. \tag{4.2}$$

We will use this to show in [CGJ] that $\theta(f)$ is not overconvergent when neither f nor k equals zero.

5. Cohomology

For a rigid analytic open subspace W of $X_1(N)$ set

$$H(k)(W) =: (\Omega^1_{X_1(N)}(\log C) \otimes \mathscr{H}_k)(W)/\nabla\mathscr{H}_k(W)\,,$$

and let $\Omega^\bullet(\mathscr{H}_k)$ be the complex $\mathscr{H}_k \xrightarrow{\nabla} \Omega^1_{X_1(N)}(\log C) \otimes \mathscr{H}_k$.

Let SS denote the set of supersingular points on $X_1(N)$ mod p and let $\widetilde{SS}$ be a set of liftings. (We will also denote by SS the degree of the divisor SS when appropriate.) By [BC, Theorems 2.1 and 2.4], via the natural maps $H(k)(W_1)$ and $H(k)(W_2)$ are both isomorphic to $\mathbf{H}^1(X_1(N), \Omega^\bullet(\mathscr{H}_k)(\log \widetilde{SS}))$ where $\Omega^\bullet(\mathscr{H}_k)(\log \widetilde{SS})$ is the complex

$$\mathscr{H}_k \xrightarrow{\nabla} \Omega^1_{X_1(N)}(\log(C \cup \widetilde{SS})) \otimes \mathscr{H}_k\,. \tag{5.1}$$

We therefore let $H(k)$ denote any one of these cohomology groups. The above, 2.2 and 2.4 imply

Theorem 5.1 *The space $H(k)$ is finite dimensional and F_k and V_k induce endomorphisms Frob and Ver of $H(k)$ such that*

$$Frob \circ Ver = Ver \circ Frob = p^{k+1}\,.$$

Let $]C[$ denote the union of the cuspidal residue classes and $]SS[= W_1 - Z$. Then $]SS[$ is the inverse image in W_1 under reduction of SS and after étale base extension is a disjoint union of wide open annuli, one for each element of SS. We call these the supersingular annuli. Now let $\Omega^\bullet(\mathscr{H}_k) \otimes \mathscr{I}_C$ be the subcomplex of $\Omega^\bullet(\mathscr{H}_k), \mathscr{H}_k \otimes \mathscr{I}_C \xrightarrow{\nabla} \Omega^1_{X_1}(N) \otimes \mathscr{H}_k$, where $\mathscr{I}_C$ is the ideal sheaf of the cusps. We let $H_{\mathrm{par}}(k)$ denote the kernel of the map $H(k) \to H(k)(]C[\cup]SS[)$. This is naturally isomorphic to the classical weight k parabolic cohomology on $X_1(N)$ which is the image of $\mathbf{H}^1(X_1(N), \Omega^\bullet(\mathscr{H}_k) \otimes \mathscr{I}_C)$ in $\mathbf{H}^1(X_1(N), \Omega^\bullet(\mathscr{H}_k))$. Also, $H_{\mathrm{par}}(k)$ is stable under *Frob* and *Ver* and

Theorem 5.2 *There is a natural perfect pairing*

$$(\,,\,)\;:\; H_{\mathrm{par}}(k) \times H_{\mathrm{par}}(k) \to K$$

such that

$$(Frob(\alpha), Frob(\beta)) = p^{k+1}(\alpha, \beta)\,.$$

Proof. It is classical that the self-duality of $\mathscr{H}_k$ leads to a perfect pairing on $H_{\mathrm{par}}(k)$ (essentially Poincaré duality between compactly and non-compactly supported cohomology). By standard arguments (e.g. see [RLC, Thm. 4.5]) we can compute it as follows. Let h and g be elements of $(\mathscr{H}_k \otimes \Omega^1_{X_1(N)})(W_1)$ with trivial residues on the supersingular annuli and let $[h]$ and $[g]$ denote their respective cohomology classes in $H_{\mathrm{par}}(k)$. It follows, in particular, that for each supersingular annulus A, there exists a $\lambda_A \in \mathscr{H}_k(A)$ such that $\nabla\lambda_A = f|_A$. For each supersingular point x of $X_1(N)$ let A_x denote the supersingular annulus above x. Then

$$([h], [g]) = \sum_{x \in SS} Res_{A_x}(\lambda_{A_x}, g|_{A_x}), \tag{5.2}$$

where Res_{A_x} is the residue map associated to the orientation on A_x coming from W_1 (see [RLC, Sect. 3]). Now let T_x be an orientation preserving uniformizing parameter on A_x. We may write

$$h|_{A_x} = \sum_n b_{x,n} T_x^n \frac{dT_x}{T_x} \quad \text{and} \quad g|_{A_x} = \sum_n c_{x,n} T_x^n \frac{dT_x}{T_x},$$

where the $b_{x,n}$, $c_{x,n} \in \mathscr{H}_k(A_x)^{\nabla}$. Then

$$(\lambda_{A_x}, g_{A_x}) = \sum_n \left(\sum_{i+j=n} (b_{x,i}, c_{x,j})/i \right) T_x^n \frac{dT_x}{T_x}$$

and $(b_{x,i}, c_{x,j}) \in K$. Also, if $f = \sum_n a_n T_x^n \, dT_x/T_x$ with $a_n \in \mathscr{H}_k(A_x)^{\nabla}$,

$$Ff = \textstyle\sum_n \pi^* a_n^{(\phi^*)} \phi^* \left(T_x^n \frac{dT_x}{T_x} \right) .$$

We deduce that

$$\begin{aligned} &(Frob([h]), Frob([g])) \\ &\quad = \sum_{x \in SS} Res_{A_{\phi^{-1}x} \cap W_2} \left(\phi^* \sum_n \left(\sum_{i+j=n} (\pi^* b_{x,i}^{(\phi^*)}, \pi^* c_{x,j}^{(\phi^*)})/i \right) T_x^n \frac{dT_x}{T_x} \right) . \end{aligned}$$

Now we can identify the space of horizontal sections of ∇ on A_x with $Sym^k H^1_{\text{Cris}}(E_x, K)$ where E_x is the supersingular elliptic curve corresponding to x, the restriction of $(\,,)$ to this space with the natural pairing, and the map from $\mathscr{H}_k(A_x)^{\nabla}$ to $\mathscr{H}_k(A_{\phi^{-1}(x)})^{\nabla}$, $a \mapsto \pi^* a^{(\phi^*)}$, with the Frobenius morphism from $Sym^k H^1_{\text{Cris}}(E_x, K)$ to $Sym^k H^1_{\text{Cris}}(E_{\phi^{-1}x}, K)$. It follows that

$$(\pi^* b_{x,i}^{(\phi^*)}, \pi^* c_{x,j}^{(\phi^*)}) = p^k (b_{x,i}, c_{x,j}) .$$

The result follows from this and the fact that $Res_{A_{\phi^{-1}x}} \phi^* g = p \, Res_{A_x} g$. $\square$

It follows immediately from Theorems 5.1 and 5.2 that

Corollary 5.2.1 *We have*

$$(Frob(\alpha), \beta) = (\alpha, Ver(\beta)) \quad \text{and} \quad (Ver(\alpha), Ver(\beta)) = p^{k+1}(\alpha, \beta) .$$

By a *generalized eigenvector* with eigenvalue $a \in K$ for a linear operator L on a vector space W over K, we mean a vector in W which is in the kernel of $(L - a)^n$ for some positive integer n.

Theorems 5.1, 5.2 and the previous corollary imply

Corollary 5.2.2 *The map Ver is an isomorphism and if* $\alpha \in K^*$ *then the dimension of the generalized eigensubspace of* $H_{\mathrm{par}}(k)$ *with eigenvalue* α *for Ver is equal to that of the generalized eigensubspace with eigenvalue* p^{k+1}/α.

Note. We have not yet shown that the eigenvalues of Ver acting on $H(k)$ are integers in $\mathbf{C}_p$. This will follow from Theorem 6.1.

Lemma 5.3

$$\dim_K H(k)/H_{\mathrm{par}}(k) = \begin{cases} C + SS - 1 & \text{if } k = 0, \\ C + (k+1)SS & \text{otherwise}. \end{cases}$$

Proof. We take cohomology of the short exact sequence,

$$0 \to \Omega^{\bullet}(\mathscr{H}_k) \otimes \mathscr{I}_C \to \Omega^{\bullet}(\mathscr{H}_k)(\log(\widetilde{SS})) \to \mathscr{S}_k^{\bullet} \to 0,$$

where $\mathscr{S}_k^{\bullet}$ is the complex of skyscraper sheaves which makes this sequence exact. First $\mathscr{S}_k^0 \cong \mathscr{H}_k/\mathscr{I}_C\mathscr{H}_k$ and

$$\mathscr{S}_k^1 \cong (\mathscr{H}_k \otimes \mathscr{O}_{X(N;p)}/\mathscr{I}_C) \otimes (\mathscr{H}_k \otimes \mathscr{O}_{X(N;p)}/\mathscr{I}_{\widetilde{SS}})$$

where $\mathscr{I}_{\widetilde{SS}}$ is the ideal sheaf of $\widetilde{SS}$ (via a residue map). Second, the boundary map of the complex $\mathscr{S}^{\bullet}$ takes $\mathscr{S}_k^0$ into $\mathscr{H}_k \otimes \mathscr{O}_{X(N;p)}/\mathscr{I}_C$ with respect to this decomposition with a one dimensional cokernel at each cusp (as one can deduce from the results of [K-pMF, A1]). Thus as $\mathscr{H}_k$ is locally free of rank $k+1$ we see that $\dim_K \mathbf{H}^1(\mathscr{S}_k^{\bullet}) = C + (k+1)SS$. The lemma follows from the fact that $\dim_K \mathbf{H}^2(\Omega^{\bullet}(\mathscr{H}_k)(\log(\widetilde{SS}))) = 0$ for $k \geqq 0$ and $\dim_K \mathbf{H}^2(\Omega^{\bullet}(\mathscr{H}_k) \otimes \mathscr{I}_C) = 0$ if $k > 0$ and 1 if $k = 0$. $\square$

Now on q-expansions it is evident that $\theta \circ U = pU \circ \theta$. Using this, Proposition 4.1 and Lemma 4.2 we deduce:

Theorem 5.4 *The quotient* $M_{k+2}/\theta^{k+1}M_{-k}$ *is naturally isomorphic to* $H(k)$ *and the following diagram commutes:*

$$\begin{array}{ccc} M_{k+2}/\theta^{k+1}M_{-k} & \xrightarrow{U} & M_{k+2}/\theta^{k+1}M_{-k} \\ \downarrow & & \downarrow \\ H(k) & \xrightarrow{Ver} & H(k). \end{array}$$

Let $P(k,T)$ denote the characteristic series of U restricted to $\omega^k(Y)$ where Y is any underlying affinoid of W_1 strictly containing Z. This series exists since U is completely continuous on this space as explained in [G-ApM]. Then,

Corollary 5.4.1

$$P(k,T) = P(2-k,\ p^{k-1}T)\det(1 - Ver\,T|H(k-2)).$$

We will interpret the polynomial $\det(1 - Ver\,T|H(k-2))$ in terms of Hecke operators and classical modular forms on $X(N;p)$ in Sect. 7. This will generalize Theorem 2 of [Ko].

6. Small slope forms are classical

For a $K[U]$ module M and a rational number α, we set M_α equal to the *slope* α part of M, that is, the submodule of $F \in M$ such that there exists a polynomial $f(T) \in K[T]$ whose roots in $\mathbf{C}_p$ all have valuation α such that $f(U)F = 0$. We say that a non-zero element of M_α has slope α.

Let $X(N;p)$ denote the fiber product of $X_1(N)$ and $X_0(p)$ over the j-line. Let $S_{k,\mathrm{cl}} := S_k(N;p) \subseteq M_{k,\mathrm{cl}} := M_k(N;p)$ denote the spaces of cusp forms and modular forms of weight k on $X(N;p)$. We may use $\mathscr{K}$ to identify W_1 with a subspace of $X(N;p)$ and get a natural Hecke compatible injection from $M_{k,\mathrm{cl}}$ into M_k under which U_p corresponds to U. We call elements in the image *classical* modular forms.

Theorem 6.1 *Every p-adic overconvergent form of weight $k+2$ and slope strictly less than $k+1$ in M_{k+2} is classical.*

An immediate consequence of this theorem and the main result of [GM-CP] is the following generalization of a result of Koike's [Ko]:

Corollary 6.1.1 *If $P_{k,\mathrm{cl}}(T) = \det(1 - U_pT|M_{k,\mathrm{cl}}(N;p))$ then if k, k' and n are integers such that $k' \geqq k \geqq 2$, $n > k-1$ and $k' \equiv k \bmod p^{n-1}(p-1)$, then*

$$P_{k,\mathrm{cl}}(T) \equiv P_{k',\mathrm{cl}}(T) \bmod p^{k-1}\,.$$

We now begin the proof of Theorem 6.1. It is vacuous for $k < 0$. Therefore suppose $k \geqq 0$. Also, suppose $N > 4$. We will explain how to handle small levels at the end of this section. Let S_k denote the subspace of cusp forms in M_k, and let S_k^0 denote the subspace of S_k of forms with trivial residues on the supersingular annuli (see [pSI]). Then $S_{k+2}^0/\theta^{k+1}M_{-k}$ is naturally isomorphic to the parabolic cohomology $H_{\mathrm{par}}(k)$. Let $S_{k,\mathrm{cl}}^0 := S_{k,\mathrm{cl}}^0(N;p)$ denote the space of p-old (or equivalently (by [pSI, Thm. 9.1]) with trivial residues on the supersingular annuli) cusp forms on $X(N;p)$ of weight k. Then $S_{k,\mathrm{cl}}^0$ maps into S_k^0 and

Lemma 6.2 *We have*

$$\dim_K(S_{k+2,\mathrm{cl}}^0) = \dim_K(H_{\mathrm{par}}(k))\,. \tag{6.1}$$

Proof. The dimensions of both spaces are twice the dimension of the space of cusp forms on $X_1(N)$. (For $H_{\mathrm{par}}(k)$, this follows from the classical Shimura isomorphism [Sh].) □

The next observation is

Lemma 6.3 *If F is a non-zero element of M_{k+2} with slope strictly less than $k+1$, it is not in $\theta^{k+1}M_{-k}$.*

Proof. We may suppose that F is an eigenform with eigenvalue γ such that $v(\gamma) < k+1$. Suppose $G \in M_{-k}$ such that $\theta^{k+1}G = F$ and let $F(q) = \sum a_nq^n$ and $G(q) = \sum b_nq^n$. Then

$$a_{pn} = \gamma a_n \quad \text{and} \quad a_n = n^{k+1}b_n\,.$$

This implies $G(q)$ has unbounded coefficients which contradicts the supposition that $G \in M_{-k}$. $\square$

We let U act on $H(k)$ as Ver.

Corollary 6.3.1 *The natural map from* $(M_{k+2,\mathrm{cl}})_\alpha$ *to* $H(k)_\alpha$ *is an injection if* $\alpha < k+1$.

In particular,

$$\dim_K (S^0_{k+2,\mathrm{cl}})_\alpha \leqq \dim_K H_{\mathrm{par}}(k)_\alpha \tag{6.2}$$

if $\alpha < k+1$. Now it follows from Corollary 5.2.2 that

$$\dim_K H_{\mathrm{par}}(k)_0 = \dim_K H_{\mathrm{par}}(k)_{k+1}\,. \tag{6.3}$$

Let f_1, f_2 denote the two degeneracy maps from $X(N; p)$ to $X_1(N)$ and let F be a form of weight $k+2$ on $X_1(N)$ so that if (E, P, C) represents a point on $X(N; p)$, where E is an elliptic curve, P is a point of order N on E and C is a subgroup of order p of E, and if $\rho_C : E \to \rho_C E$ is the isogeny with kernel C then

$$f_1^* F(E, P, C) = F(E, P)\,,$$

$$f_2^* F(E, P, C) = \rho_C^* F(\rho_C E, \rho_C P)\,.$$

Consider the identities

$$(f_1^* F)|U_p = f_1^*(F|T_p) - p^{-1} f_2^* F\,,$$

$$(f_2^* F)|U_p = p^{k+2} f_1^* F|\langle p\rangle_N\,.$$

Suppose

$$F|T_p = A_p F \quad \text{and} \quad F|\langle d\rangle = \varepsilon(d)F\,, \tag{6.4}$$

for $d \in (\mathbf{Z}/N\mathbf{Z})^*$. Let u be a root of $x^2 - A_p x + \varepsilon(p)p^{k+1}$. Then the above identities imply that $G = f_1^* F - (pu)^{-1} f_2^* F$ is an eigenvector for U_p with eigenvalue u. This implies that for any $A \in \overline{K}^*$, any character $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \overline{K}^*$ and any root u of $x^2 - Ax + \varepsilon(p)p^{k+1}$, we have a homomorphism h_u from the subspace $V(A, \varepsilon)$ of $S_{k+2}(N)$ which is the eigenspace for T_p with eigenvalue A and character ε for the action $(\mathbf{Z}/N\mathbf{Z})^*$ to the subspace $W(u, \varepsilon)$ of $S^0_{k+2,\mathrm{cl}}$ with eigenvalue u for U_p and character ε.

Lemma 6.4 *The homomorphism* h_u *is an isomorphism. Moreover*, $W(u, \varepsilon)$ *is the kernel of* $(U_p - u)^2$ *in the* ε *eigencomponent unless* $V(A, \varepsilon) \neq 0$ *and* $u = \varepsilon(p)p^{k+1}/u =: u'$. *In this case, the kernel of* $(U_p - u)^2$ *in the* ε *eigencomponent equals the kernel of* $(U_p - u)^3$ *in this component and is strictly bigger than* $W(u, \varepsilon)$.

Proof. We will use the fact that $f_1^* S_{k+2}(N) \cap f_2^* S_{k+2}(N) = \{0\}$. This already implies that h_u is an injection. Suppose $L = f_1^* H + f_2^* G \in W(u, \varepsilon)$. Then we see that

$$H|\langle d\rangle_N = \varepsilon(d)H, \quad uH = H|T_p + p^{k+2}\varepsilon(p)G \quad \text{and} \quad uG = -p^{-1}H\,.$$

This implies that $H|T_p = (u + \varepsilon(p)p^{k+1}/u)H = AH$ and $h_u H = L$. Thus $H \in V(A, \varepsilon)$ and h_u is an isomorphism. Now suppose $L|(U_p - u)^2 = 0$ and L is in the ε eigencomponent. Then by what we now know there exists an $F \in V(A, \varepsilon)$ such that

$$L|(U_p - u) = f_1^* F - (pu)^{-1} f_2^* F .$$

This implies

$$H|T_p - uH + p^{k+2}\varepsilon(p)G = F \quad \text{and} \quad -p^{-1}H + (pu)^{-1}F = uG .$$

Hence,

$$H|T_p - AH = \left(1 - \frac{u'}{u}\right) F .$$

But since T_p is diagonalizable, we must have either $F = 0$ or $u' = u$.

Finally, if $u = u'$ the above implies that if $H \in V(A, \varepsilon)$, $f_1^* H|(U_p - u) = h_u(uH)$. So if $H \neq 0$, $f_1^* H$ is in the kernel of $(U_p - u)^2$ but is not in $W(u, \varepsilon)$. $\square$

Corollary 6.4.1 *Suppose* $\alpha \leqq (k+1)/2$. *Then*

$$\dim_K (S^0_{k+2,\mathrm{cl}})_\alpha = \dim_K (S^0_{k+2,\mathrm{cl}})_{k+1-\alpha} , \tag{6.5}$$

and

$$(S^0_{k+2,\mathrm{cl}})_\alpha + (S^0_{k+2,\mathrm{cl}})_{k+1-\alpha} = f_1^* S_{k+2}(\Gamma_1(N))_\alpha + f_2^* S_{k+2}(\Gamma_1(N))_\alpha ,$$

where $S_{k+2}(\Gamma_1(N))_\alpha$ *is the slope* α *subspace of* $S_{k+2}(\Gamma_1(N))$ *for* T_p *when* $\alpha < (k+1)/2$ *and is the sum of the subspaces of slope at least* $(k+1)/2$ *when* $\alpha = (k+1)/2$.

Proof. Using the fact that T_p is diagonizable on $S_{k+2}(N)$, the lemma implies that if $\alpha < (k+1)/2$, both the dimensions in (6.5) equal $\dim_K(S_{k+2}(\Gamma_1(N))_\alpha)$. The second statement is an immediate consequence of the lemma if $\alpha < (k+1)/2$ and is a consequence of the proof if $\alpha = (k+1)/2$. $\square$

Putting (6.1), (6.2), (6.3) and (6.5) together we deduce all the inequalities in (6.2) are equalities and so

Lemma 6.5 *The map from* $(S^0_{k+2,\mathrm{cl}})_\alpha$ *to* $H_{\mathrm{par}}(k)_\alpha$ *is an isomorphism if* $\alpha < k+1$.

This is enough to prove Theorem 6.1 for elements of S^0_{k+2}, as we shall see. To deal with arbitrary elements of M_{k+2} we will need

Proposition 6.6 *The map from* $(M_{k+2,\mathrm{cl}})_\alpha$ *to* $H(k)_\alpha$ *is an isomorphism if* $\alpha < k+1$.

Proof. First we have

Lemma 6.7 *Suppose* $\alpha \leqq k+1$. *Then*

$$\dim_K (M_{k+2,\mathrm{cl}})_\alpha/(S^0_{k+2,\mathrm{cl}})_\alpha = \begin{cases} C + SS - 1 & \text{if } \alpha = k/2 = 0, \\ C & \text{if } \alpha = 0 \text{ or } k+1 \text{ and } k > 0, \\ C - 1 & \text{if } \alpha = k+1 = 1, \\ (k+1)SS & \text{if } \alpha = k/2 \text{ and } k > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Let $E_k(N)$ and $E_k(N;p)$ denote the spaces of Eisenstein series of weight k over K on $X_1(N)$ and on $X(N;p)$ and let $S^{\mathrm{new}}_{k,\mathrm{cl}} =: S^{\mathrm{new}}_k(N;p)$ denote the space of cusp forms new at p on $X(N;p)$ identified with its image in $M_k(\Gamma_1(N))$. First the map from $E_{k+2}(N;p) + S^{\mathrm{new}}_{k+2,\mathrm{cl}}$ to $M_{k+2,\mathrm{cl}}/S^0_{k+2,\mathrm{cl}}$ is an isomorphism. Now $S^{\mathrm{new}}_{k+2,\mathrm{cl}} \subseteq (M_{k+2,\mathrm{cl}})_{k/2}$ and by [pSI, Lemma 5.1] has dimension $SS(k+1)$ unless $k=0$ in which case it has dimension $SS-1$. Also $E_k(N;p) \subseteq (M_{k+2,\mathrm{cl}})_0 + (M_{k+2,\mathrm{cl}})_{k+1}$ and $S^{\mathrm{new}}_{k+2,\mathrm{cl}} \cap E_{k+2}(N;p) = 0$. From this we deduce the lemma when $\alpha \neq 0$ or $k+1$. If $F \in E_{k+2}(N)$ is an eigenform for T_p with eigencharacter ε, its eigenvalue is $\varepsilon_1 + \varepsilon_2 p^{k+1}$ where ε_1 and ε_2 are roots of unity of relatively prime order such that $\varepsilon_1\varepsilon_2 = \varepsilon(p)$ (see [FJ, Sect. 3]). It follows that

$$F_1 = f_1^* F - (\varepsilon_1 p)^{-1} f_2^* F \quad \text{and} \quad F_2 = f_1^* F - (\varepsilon_2 p^{k+2})^{-1} f_2^* F \tag{6.6}$$

are eigenvectors for U_p in $E_{k+2}(N;p)$ with eigenvalues ε_1 and $\varepsilon_2 p^{k+1}$. Since T_p is diagonizable on $E_{k+2}(N)$ if $k \geqq 0$, $\dim_K E_{k+2}(N) = C$ if $k > 0$ and $\dim_K E_{k+2}(N;p) = 2C$ if $k > 0$ the lemma follows as long as $k > 0$.

Now suppose $k = 0$. Then by the above arguments since $\dim_K E_2(N) = C - 1$, $\dim_K E_2(N;p)_\alpha \geqq C - 1$ if $\alpha = 0$ or 1. However the pullback P of the one dimensional space of Eisenstein series on $X_0(p)$ to $X(N;p)$ lies in $E_2(N;p)_0$ because U_p on $X_0(p)$ acts on weight 2 forms as minus the Atkin–Lehner involution. On the other hand, this space is not contained in $f_1^*E_2(N) + f_2^*E_2(N)$. This can be seen by considering weight 2 forms as differentials. Then, if C_∞ is the set of cusps on $X(N;p)$ lying over ∞ on $X_0(p)$, $\sum_{P \in C_\infty} Res P\omega$ equals zero if $\omega \in f_1^*E_2(N) + f_2^*E_2(N)$ but not if $\omega \in P$. Since $\dim_K E_2(N;p) = 2C - 1$, this implies the remaining cases of the lemma. □

Now consider the diagram with exact rows:

$$\begin{array}{ccccccccc} 0 & \to & (S^0_{k+2,\mathrm{cl}})_\alpha & \to & (M_{k+2,\mathrm{cl}})_\alpha & \to & (M_{k+2,\mathrm{cl}})_\alpha/(S^0_{k+2,\mathrm{cl}})_\alpha & \to & 0\,, \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \to & H_{\mathrm{par}}(k)_\alpha & \to & H(k)_\alpha & \to & H(k)_\alpha/H_{\mathrm{par}}(k)_\alpha & \to & 0\,. \end{array} \tag{6.7}$$

Since, for $\alpha < k+1$, the first vertical arrow is an isomorphism and the second is an injection the third is an injection as well. Thus the map

$$((M_{k+2,\mathrm{cl}})_0 + (M_{k+2,\mathrm{cl}})_{k/2})/((S^0_{k+2,\mathrm{cl}})_0 + (S^0_{k+2,\mathrm{cl}})_{k/2}) \to H(k)/H_{\mathrm{par}}(k)$$

is an injection. By Lemmas 5.3 and 6.7 both these vector spaces have the same dimension so this map is an isomorphism. It follows that the last vertical arrow in (6.7) is an isomorphism and Proposition 6.6 follows from this and Lemma 6.5. $\square$

Since the eigenvalues of U_p acting on classical modular forms are algebraic integers we deduce

Corollary 6.7.1 *The eigenvalues of Frob and Ver acting on $H(k)$ are algebraic integers.*

Finally suppose G is a generalized eigenform for U in M_{k+2} of slope $\alpha < k+1$. Then, by Proposition 6.6, its class in $H(k)$ is equal to the class of a classical generalized eigenform F for U_p of slope α. Hence $G-F \in (M_{k+2})_\alpha$ and its class in $M_{k+2}/\theta^{k+1}M_k$ is 0. By Lemma 6.3 we see that $G-F=0$ which proves Theorem 6.1. $\square$

Remark. We can make Theorem 6.1 work for small levels.

Suppose first $N > 4$, $(R,p)=1$ and $N|R$. The map from $M_k(\Gamma_1(N))$ to $M_k(\Gamma_1(R))$ is clearly a U equivariant injection (look at q-expansions). We will now choose R large enough so that all the levels mentioned in this paragraph divide R and identify $M_k(\Gamma_1(N))$ with its image in $M_k(\Gamma_1(R))$. One can show (again using q-expansions) that if $(M,N) > 4$,

$$M_k(\Gamma_1((M,N))) = M_k(\Gamma_1(M)) \cap M_k(\Gamma_1(N)) \, .$$

Now suppose $N \leqq 4$, $A,B > 4$, $(AB,p)=1$ and $(A,B)=N$. Then, it follows from the above that the intersection of $M_k(\Gamma_1(A))$ and $M_k(\Gamma_1(B))$ is independent of A and B and is stable under U. All this is compatible with changing R. We set $(M_k(\Gamma_1(N)),U)$ equal to any element in this isomorphism class of U modules. One can show this is compatible with Katz's definition of overconvergent forms in small level. Since classical forms which are both of level $\Gamma_1(A)$ and of level $\Gamma_1(B)$ are of level $\Gamma_1((A,B))$ it follows from Theorem 6.1 that its statement is now also true for all $N \geqq 1$.

Note. 1. Since f_2^*/p is Frobenius, it follows that

$$Frob(Fil^{k+1}(H_{\mathrm{par}}(k))_{>0} \cap Fil^{k+1}(H_{\mathrm{par}}(k))) = 0 \, .$$

2. The image of the form F_2 which has slope $k+1$ in (6.6) in $H(k)$ is 0 using the fact that f_2^*/p is Frobenius and thus there must be an overconvergent modular form H of weight $-k$ and slope 0 such that $\theta^{k+1}H = F_2$. In fact, if $F = G_{k+2}$, $k > 0$, the classical Eisenstein series of level one and weight $k+2$, H is the p-adic Eisenstein series G^*_{-k} (see [S, Sect. 1.6]).

7. The boundary case

We now know that no overconvergent forms of weight k of slope strictly larger than $k-1$ are classical and all of strictly smaller slope are. In this section, we will investigate the boundary case of forms of weight k and slope $k-1$.

Suppose now $k \geqq 2$. It follows from the results of Sect. 6 that

$$\dim_K (S_{k,\mathrm{cl}})_{k-1} = \dim_K\ H(k-2)_{k-1}$$

for $k \geqq 2$ (we will prove subsequently that these two spaces are isomorphic Hecke modules). Hence, it will follow from the following proposition and Note 2 of Sect. 6 that there exist non-classical overconvergent modular forms of weight k and slope $k-1$.

Suppose L is an imaginary quadratic field of discriminant D, $\sigma : L \to \mathbf{C}$ is an embedding and ψ is a Grössencharacter of L with infinity type σ^{k-1} and conductor $\mathscr{M}$. Let ϕ be the Dirichlet character associated to L, M the norm $N(\mathscr{M})$ of $\mathscr{M}$ and ε the Dirichlet character modulo $|D|M$ given by the formula

$$\varepsilon(a) = \phi(a)\psi(a)/\sigma(a)^{k-1} .$$

Then there exists a weight k cuspidal newform on $X_1(|D|M)$, G_ψ, with character ε and q-expansion

$$\sum_{\mathscr{A}} \psi(\mathscr{A}) q^{N(\mathscr{A})}$$

where the sum is over integral ideals $\mathscr{A}$ prime to $\mathscr{M}$ (see [R, Sh-CM]). (This is true even if $k = 1$ as long as ψ is not the composition of a Dirichlet character with the norm (see [M, Thm. 4.8.2]).)

Now, identify $\mathbf{C}$ with $\mathbf{C}_p$. Suppose p splits in L, $DM|N$ and let $\wp$ be the prime of L above p such that $v(\psi(\wp)) = k-1$. Then the coefficient of q^p in $G_\psi(q)$ is $\psi(\wp) + \psi(\bar{\wp})$. It follows that $F_\psi = f_1^* G_\psi - (p \quad (\wp))^{-1} f_2^* G_\psi$ is an eigenform for U_p on $X(N;p)$ with eigenvalue $\psi(\wp)$ and so lies in $(S_{k,\mathrm{cl}})_{k-1}$.

Proposition 7.1 *The image of F_ψ in $H(k-2)$ is zero.*

Proof. First, we observe that

$$F_\psi(q) = \sum_{\wp \not| \mathscr{A}} \psi(\mathscr{A})\, q^{N(\mathscr{A})} . \tag{7.1}$$

Now choose a sequence of positive integers $\{r_n\}$ such that $\lim_{n\to\infty} r_n = -1$ where the limit is taken in $\hat{\mathbf{Z}}$. Then the sequence $G_{\bar{\psi}^{r_n}}$ converges in the sense of [S] to the weight $2-k$ modular form H of slope 0 with q-expansion

$$\sum_{\wp \not| \mathscr{A}} \bar{\psi}(\mathscr{A})^{-1} q^{N(\mathscr{A})} .$$

Since H has slope zero it lies in M_{2-k} [G-ApM, Prop. II.3.22]. As $N(\mathscr{A})^{k-1} = \psi(\mathscr{A})\bar{\psi}(\mathscr{A})$ it follows from (7.1) that $\theta^{k-1} H = F$. This completes the proof. □

Remarks. 1. Suppose $G(q) = \sum_{n=1} A_n q^n$ is the normalized weight 2 cusp form on $X_0(49)$ and p is a non-zero square modulo 7. Then A_p is a unit modulo p. Using the notation of (6.4) and subsequent lines, if u is the non-unit root of $x^2 - A_p x + p$ and $F = f_1^* G - (pu)^{-1} f_2^* G$, then since G has CM by $\mathbf{Z}[\sqrt{-7}]$ the previous proposition implies that the image of F in $H(0)$ is zero.

2. It would be interesting to know whether there are any non-CM classical weight k cusp forms whose image in $H(k-2)$ is zero.

3. Another way to obtain non-classical forms of slope $k-1$ and weight k is to use the map θ^{k-1}. Indeed,

$$\theta^{k-1}(M_{2-k})_0 \subseteq (S_k)_{k-1} = (S_k^0)_{k-1} .$$

So for example, by standard procedure we may deal with the λ-line (which sits between $X_0(4)$ and $X_1(4)$). Then $(S_{2,\mathrm{cl}})_1 = 0$ and $\dim_K(S_0)_0 = (p-1)/2$ by [D]. Hence, in this case, $(S_2)_1$ is a $(p-1)/2$ dimensional space of non-classical overconvergent modular forms of weight 2 and slope 1. (See also [A] where the dimension of the space of overconvergent forms of weight 0 and slope 0 of level 3 is computed when $p \equiv 1 \bmod 3$.)

4. In particular, if $p = 13$ and A is Atkin's overconvergent weight 0 form of level 1 (see [K-pMF], Sect. 3.12]), θA is a non-classical overconvergent form of level 1, weight 2 and slope 1.

Let $\mathbf{T}$ be the free K-algebra generated by the symbols A_l, $l \neq p$ and $\langle d\rangle_N$, $d \in (\mathbf{Z}/N\mathbf{Z})^*$ modulo the relations $\langle d\rangle_N\langle d'\rangle_N = \langle dd'\rangle_N$. Then both $H_{\mathrm{par}}(k-2)$ and $S^0_{k,\mathrm{cl}}$ are naturally $\mathbf{T}[U]$ modules via the homomorphism which sends A_l to T_l if $l \nmid Np$ and U_l if $l|N$, U to U_p and $\langle d\rangle_N$ to $\langle d\rangle_N$. Moreover, the natural map

$$\iota: \ S^0_{k,\mathrm{cl}} \to H_{\mathrm{par}}(k-2)$$

is compatible with this action, but as we have seen it is not generally an isomorphism even though the two spaces have the same dimension. However,

Theorem 7.2 *The two* $\mathbf{T}[U]$*-modules* $S^0_{k,\mathrm{cl}}$ *and* $H_{\mathrm{par}}(k-2)$ *are isomorphic.*

Proof. This will follow from Lemma 6.5 and

Lemma 7.3 *There exists a non-degenerate* (*and non-canonical*) $\mathbf{T}[U]$*-equivariant pairing*

$$[\,,]\,:\, H_{\mathrm{par}}(k-2)_{k-1} \times (S^0_{k,\mathrm{cl}})_{k-1} \to K\ .$$

Proof. We regard points on $X(N,p)$ as triples (E,P,C) where E is an elliptic curve, P is a point of order N on E and C is a subgroup scheme of rank p of E. Fix a primitive Nth root of unity ζ_N in $\overline{K}$. Let w_p and w_N be the automorphisms of $X(N;p)$:

$$w_p : (E,P,C) \mapsto (\rho_C E, \rho_C P, \rho_C(E[p]))\ ,$$

$$w_N : (E,P,C) \mapsto (\rho_{(P)}E, Q, \rho_{(P)}C)\ ,$$

where (P) is the subgroup scheme generated by P, $\rho_A : E \to \rho_A E$ is the isogeny with kernel A and $Q = \rho_{(P)}(Q')$ where Q' is a point of order N on E which satisfies $(P,Q')_{\mathrm{Weil}} = \zeta_N$. Then, if F is a modular form of weight k on $X_1(N)$,

$$(f_1^*F)|w_p = f_2^*F$$

and

$$(f_2^* F)|w_p = p^k f_1^* F|\langle p \rangle_N \ .$$

First, the operator U is invertible on both $H_{\mathrm{par}}(k)$ and $S^0_{k,\mathrm{cl}}$ and the pairing $(\,,\,)_k$ on $H_{\mathrm{par}}(k-2)$ satisfies

$$(m|h, n)_k = (m, n|(w_N^{-1} h w_N))_k \tag{7.2a}$$

for $h \in \mathbf{T}$ and

$$(m|U, n)_k = (m, n|(p^{k-1}/U))_k \ . \tag{7.2b}$$

Next consider the map $r : S^0_{k,\mathrm{cl}} \to S^0_{k,\mathrm{cl}}$ defined by

$$r(G) = G| \left((p-1)U + (1 - U^2/\langle p \rangle p^{k-1}) w_p \right) \ .$$

This map satisfies

$$r \circ h = h \circ r \ ,$$

$$r \circ U = (\langle p \rangle_N p^{k-1}/U)) \circ r \tag{7.3}$$

for $h \in \mathbf{T}$. In fact, $r \circ h_u = (pu - u')h_{u'}$ on $V(A, \varepsilon)$ where u and u' are the two roots of $x^2 - Ax + \varepsilon(p)p^{k-1}$. In particular, the restriction of r to $(S^0_{k,\mathrm{cl}})_{k-1}$ is an isomorphism onto $(S^0_{k,\mathrm{cl}})_0$. We may now set

$$[m, n] = (m, \imath(r(n))|w_N)_k$$

for $m \in H_{\mathrm{par}}(k-2)_{k-1}$ and $n \in (S_{k,\mathrm{cl}})_{k-1}$. That this pairing is $\mathbf{T}[U]$-equivariant follows from (7.2) and (7.3). That it is non-degenerate follows from the fact that the restriction of $\imath$ to $(S^0_{k,\mathrm{cl}})_0$ is an injection onto $H_{\mathrm{par}}(k-2)_0$. □

We do get the following positive result in the boundary case,

Corollary 7.2.1 *Suppose $k \geqq 2$. If F is an overconvergent weight k Hecke eigenform of slope $k-1$ such that $F \notin \theta^{k-1} M_{2-k}$, then F is classical.*

Proof. Since F is assumed to have positive slope it must be cuspidal, the image of F in $H(k-2)$ is non-trivial and lies in $H_{\mathrm{par}}(k-2)$. Let $F(q) = \sum_{n=1}^{\infty} a_n q^n$. The theorem implies that there exists a classical form G such that $G|A_l = a_l G$ for l a prime not equal to p and $G|U = a_p G$. Hence, G has the same q-expansion as F and so the two forms are equal. □

Remark. Richard Taylor has asked whether an overconvergent eigenform of weight 1 and slope zero is classical if the image of inertia at p under the representation associated to F is finite.

Let $h(i) : \mathbf{T}[U] \to \mathbf{T}[U]$ denote the homomorphism such that

$$h(i)(A_l) = l^i A_l, \quad h(i)(\langle d \rangle_N) = \langle d \rangle_N \quad \text{and} \quad h(i)(U) = p^i U \ .$$

For a $\mathbf{T}[U]$-module M and an integer i, we define the twisted $\mathbf{T}[U]$ module $M(i) = M \otimes_{h(i)} \mathbf{T}[U]$.

Corollary 7.2.2 *The* $\mathbf{T}[U]$ *module* $(M_k)_{k-1}$ *sits in an exact sequence*

$$0 \to ((M_{2-k})_0/A)(k-1) \to (M_k)_{k-1} \to (S_{k,\mathrm{cl}})_{k-1} \to 0 \ ,$$

where $A = K$ *if* $k = 2$ *and* $A = 0$ *otherwise.*

Proof. This follows immediately from the theorem and the fact that

$$\theta^{k-1} \circ h(k-1)(g) = g \circ \theta^{k-1}$$

for $g \in \mathbf{T}[U]$. □

It follows from [H] and the results of Sect. 6 that

Corollary 7.2.3 *Suppose* $k' \equiv 2-k \mod (p-1)$ *and* $k' \geqq 2$. *Then*

$$\dim_K(M_k)_{k-1} = \dim_K(S_{k,\mathrm{cl}})_0 + \dim_K(M_{k',\mathrm{cl}})_0 - \delta \ ,$$

where δ *is SS if* $k = 2$ *and* 0 *otherwise.*

Remark. Let F and notation be as in Remark 1 above. Let m_F be the maximal ideal of $\mathbf{T}[U]$ corresponding to F. It follows from Theorem 7.2 that U does not act semi-simply on the m_F-isotypic component of $S_2(\Gamma_0(49))$. Thus there exists an overconvergent form $H \in S_2(\Gamma_0(49))_1$ which is not an eigenform and p-adic numbers m_l for each prime number l such that

$$H|T_l - A_l H = m_l F$$

for prime $l \neq 7$ and

$$H|U - uH = m_p F \ .$$

At present we have no further information about H. We do not even know whether H is an element of $\theta M_0(\Gamma_0(49))$. More precisely, Theorem 7.2 tells us that the dimension of the m_F-isotypic component of $S_2(\Gamma_0(49))$ is at least 2 and the previous corollary tells us that it is at most $\dim_K(M_{p-1,\mathrm{cl}})_0$ but we do not know what it is.

In summary, the main results of this section have concerned relationships among three spaces of overconvergent modular forms; $(M_k)_{k-1}$ and its two subspaces $(S_k)_{k-1}$ and $\theta^{k-1}(M_{2-k})_0$. We have shown that $(M_k)_{k-1}/\theta^{k-1}(M_{2-k})_0$ is isomorphic to $(S_k)_{k-1}$ as a Hecke module (Theorem 7.2) but that $\theta^{k-1}(M_{2-k})_0 \cap (S_k)_{k-1}$ may be non-zero (Theorem 7.1) and it is if and only if there exist non-classical forms in $(M_k)_{k-1} - \theta^{k-1}(M_{2-k})_0$. Moreover, none of the latter can be eigenforms for the full Hecke algebra (Corollary 7.2.1).

We also deduce from Proposition 6.6 and Theorem 7.2

Proposition 7.4

$$\det(1 - \mathit{Ver}\, T|H(\Gamma_1(N))(k-2)) = \frac{\det(1 - U_p T|M_k(N;p))}{\det(1 - U_p T|E_k(N;p)_{k-1})} \ .$$

We can relate the right hand side of the above expression to the classical Hecke polynomials [I]. *When* $k > 2$,

$$\frac{\det(1 - U_pT|M_k(N;p))}{\det(1 - U_pT|S_k^{\rm new}(N;p))} = \det(1 - T_pT + \langle p\rangle p^{k-1}T^2|\mathscr{M}_k(\Gamma_1(N)))\ ,$$

where $\mathscr{M}_k(\Gamma_1(N))$ *denotes the space of modular forms of weight* k *on* $X_1(N)$. *When* $k = 2$ *this formula holds with the left hand side divided by* $1 - T$.

8. Level *Np*

In this section we will explain how to generalize the results of the previous sections to modular forms on $X_1(Np)$. We could have worked in this generality from the beginning but thought the extra complications would have been too distracting.

Suppose $\mu_p \subset K$. As explained in Sect. 6, the subgroup $\mathscr{K}$ of E_1 gives us an embedding of W_1 into $X(N;p)$ so that E_1 is the pullback of the universal elliptic curve over $X(N;p)$ with $\Gamma_1(N)\cap\Gamma_0(p)$ structure. We will henceforth regard Z and W_i as rigid subspaces of $X(N;p)$. Let $g : X_1(Np) \to X(N;p)$ be the natural map, $Z(p) =: Z_1(Np) =: g^{-1}Z$ and $W_i(p) = g^{-1}(W_i)$. Also let $f_{(p)} : E_1(Np) \to X_1(Np)$ be the universal generalized elliptic curve over $X_1(Np)$ and $E_i(p) = E_1(Np)|_{W_i}$. Then we have a commutative diagram analogous to (2.1).

We can define sheaves analogous to ω and sheaves with connection analogous to $(\mathscr{H}_k,\ \nabla_k)$ on $X_1(Np)$ by the same procedures followed in Sect. 2 using $E_1(Np)$ in place of $E_1(N)$ (or we can just pull back the ones we have from $X_1(N)$) and we will denote them by $\omega(p)$ and $(\mathscr{H}_k(p),\ \nabla_k(p))$, respectively. We call sections of ω^k on $Z_1(Np)$ *convergent forms* of level $\Gamma_1(Np)$, sections on any strict neighborhood of $Z_1(Np)$ *overconvergent forms* of level $\Gamma_1(Np)$ and we set $M_k(p) = M_k(\Gamma_1(Np)) =: \omega^k(W_1(p))$. Then we can define a U operator on $M_k(p)$ in the same way as before.

Restriction gives a natural map from $M_{k,\rm cl}(p)$, the space of weight k modular forms on $X_1(Np)$, into $M_k(p)$ and we call the elements in the image *classical*. We have the following generalization of Theorem 6.1.

Theorem 8.1 *Every* p*-adic overconvergent form of weight* $k + 2$ *and slope strictly less than* $k + 1$ *in* $M_{k+2}(p)$ *is classical.*

The definitions and results of Sects. 2–4 carry over without difficulty. For a rigid open subspace W of $X_1(Np)$ set

$$H(k,p)(W) = (\Omega^1_{X_1(Np)}(\log\ C(p)\otimes\mathscr{H}_k(p))(W))/\nabla_k(p)\mathscr{H}_k(p)(W)\ ,$$

let $]C(p)[$ denote the union of the cuspidal residue classes, $]SS(p)[= W_1(p) - Z(p)$ and let $H_{\rm par}(k,p)(W)$ denote the kernel of the map

$$H(k,p)(W) \to H(k,p)\ ((]C(p)[\cup]SS(p)[)\cap W)\ .$$

The only results of Sect. 5 which require additional comment in this context are that $H(k, p)(W_1(p))$ is finite dimensional, the natural map from $H(k, p)(W_1(p))$ to $H(k, p)(W_2(p))$ is an isomorphism and $H_{\mathrm{par}}(k, p)(W_1(P))$ has a natural non-degenerate pairing.

We will again use the main results of [BC], the only problem is that $W_1(p)$ is not naturally a wide open in a complete curve with good reduction to which $(\mathscr{H}_k(p),\ \nabla_k(p))$ extends. However, we can overcome this difficulty as follows.

First let W_∞ and W_0 be the inverse images of the two components of the reduction of the Deligne–Rapoport model of $X_1(Np)$ over R and suppose the cusp c is a point of W_c. Then $W_1(p) \subset W_\infty$ and the rigid space $W_\infty \cap W_0$ is a disjoint union of wide open annuli, $A_x(p)$, one for each supersingular point $x \in SS$. We may glue a disk D_x to each such annulus as in [IP, Sect. A2] to obtain a complete curve Y whose reduction is the Igusa curve of level N. Now we will define an extension of $(\mathscr{H}_k(p)|W_{1(p)},\ \nabla_k(p))$ to a sheaf $\mathscr{G}_k$ with connection on Y. Let B_x denote a basis of horizontal sections of $\nabla_x(p)$ on $A_x(p)$ for each $x \in SS$. Then $\mathscr{G}_k$ is determined by the data

$$\mathscr{G}_k(W_1(p)) = \mathscr{H}_k(p)(W_1(p)), \quad \mathscr{G}_k(D_x) = Maps(B_x,\ \mathscr{O}_Y^{\mathrm{an}}(D_x))$$

and if $f \in Maps(B_x,\ \mathscr{O}_Y^{\mathrm{an}}(D_x))$

$$Res^{D_x}_{A_x(p)} f = \sum_{B_x} f(b)b\ .$$

The extension of $\nabla_k(p)$ is determined by requiring the elements of $Maps(B_x, K)$ in $\mathscr{G}_k(D_x)$ to be horizontal. Let R be an effective divisor on Y such that there is one point (counting multiplicity) in the support of R in each disk D_x. Now we can apply [BC, Theorems 2.1 and 2.4] to conclude that the natural maps from $H(k, p)(W_\infty)$ to $H(k, p)(W_1(p))$ and from $H(k, p)(W_1(p))$ to $H(k, p)(W_2(p))$ are isomorphisms. Moreover, each of these groups is isomorphic to the first hypercohomology group of the complex

$$\Omega^{\bullet}(\mathscr{G}_k)(\log R) := \mathscr{G}_k \to \Omega^1_Y(\log(C \cup R)) \otimes \mathscr{G}_k\ .$$

Thus the space $H(k, p) := H(k, p)(W_1(p))$ is finite dimensional.

Now let P_k denote $H_{\mathrm{par}}(k, p)(X_1(Np))$. Then a Meyer–Vieitoris argument making use of [B1, B2] and the covering $\{W_\infty,\ W_0\}$ yields

$$P_k \cong H_{\mathrm{par}}(k, p)(W_\infty) \oplus H_{\mathrm{par}}(k, p)(W_0)\ . \tag{8.0}$$

In particular, the dimension over K of $H(k, p)(W_\infty)$ is finite and this can be used to give another proof of the finite dimensionality of $H(k, p)$. Moreover, the self-duality of $\mathscr{H}_k(p)$ induces a perfect pairing $(\ ,\ )_k$ on P_k and we can define pairings $(\ ,\ )_k^\infty$ and $(\ ,\ )_k^0$ on $H(k, p)(W_\infty)$ and $H(k, p)(W_0)$ by the same procedure as in (5.2) of the proof of Theorem 5.2 and show

$$(\alpha, \beta)_k = (\alpha|_{W_\infty}, \beta|_{W_\infty})_k^\infty + (\alpha|_{W_0}, \beta|_{W_0})_k^0\ ,$$

where α and β are elements of P_k. It follows that $(\ ,\ )_k^\infty$ is non-degenerate.

We also conclude in the same way as in Sect. 5 that

Proposition 8.2 *There exists an endomorphism Ver of* $H(k,p)$, *the quotient* $M_{k+2}(p)/\theta^{k+1}M_k(p)$ *is naturally isomorphic to* $H(k,p)$ *and the following diagram commutes*:

$$\begin{array}{ccc} M_{k+2}(p)/\theta^{k+1}M_{-k}(p) & \xrightarrow{U} & M_{k+2}(p)/\theta^{k+1}M_{-k}(p) \\ \downarrow & & \downarrow \\ H(k,p) & \xrightarrow{Ver} & H(k,p)\,. \end{array}$$

Moreover, if $H_{\mathrm{par}}(k,p) = H_{\mathrm{par}}(k,p)(W_1(p))$,

Proposition 8.3 *The space* $H_{\mathrm{par}}(k,p)$ *is stable under Ver and if* α *is a rational number*

$$\dim_K H_{\mathrm{par}}(k,p)_\alpha = \dim_K H_{\mathrm{par}}(k,p)_{k+1-\alpha}\,. \tag{8.1}$$

One thing we can now use is the action of $(\mathbf{Z}/p\mathbf{Z})^*$, $F \mapsto F|\langle d\rangle_p$ for $d \in (\mathbf{Z}/p\mathbf{Z})^*$, on modular forms in $M_{k,\mathrm{cl}}(p)$ and in $M_k(p)$. It also acts on $P_k, H(k,p)$ and preserves $H_{\mathrm{par}}(k,p)$. If V is any of these spaces, we set V^{new} equal to the subspace of $v \in V$ such that $\sum_{d\in(\mathbf{Z}/p\mathbf{Z})^*} v|\langle d\rangle_p = 0$. Let $S_{k,\mathrm{cl}}(p)$ denote the cusp forms in $M_{k,\mathrm{cl}}(p)$ and $S^0_{k,\mathrm{cl}}(p)$ those with trivial residues on $]SS(p)[$. It follows, in particular, that after suitable identifications

$$S^0_{k,\mathrm{cl}}(p) = S^0_{k,\mathrm{cl}} \oplus S^{\mathrm{new}}_{k,\mathrm{cl}}(p) \tag{8.2}$$

and

$$H_{\mathrm{par}}(k,p) = H_{\mathrm{par}}(k) \oplus H^{\mathrm{new}}_{\mathrm{par}}(k,p)\,. \tag{8.3}$$

Proposition 8.4 *We have*

$$\dim_K(S^0_{k+2,\mathrm{cl}}(p)) = \dim_K(H_{\mathrm{par}}(k,p))\,. \tag{8.4}$$

Proof. In view of (8.2) and (8.3) and using Lemma 6.2, we only have to show $\dim_K(S^{\mathrm{new}}_{k+2,\mathrm{cl}}(p)) = \dim_K(H^{\mathrm{new}}_{\mathrm{par}}(k,p))$. On the one hand, the classical Shimura isomorphism tells us that

$$\dim_K P^{\mathrm{new}}_k = 2\dim_k S^{\mathrm{new}}_{k+2,\mathrm{cl}}\,. \tag{8.5}$$

On the other hand, a Meyer–Vieitoris argument, as above, applied to the covering $\{W_\infty, W_0\}$ of $X_1(Np)$ tells us that P^{new}_k is naturally isomorphic to

$$H^{\mathrm{new}}_{\mathrm{par}}(k,p)(W_\infty) \oplus H^{\mathrm{new}}_{\mathrm{par}}(k,p)(W_0)\,.$$

Let w be an Atkin–Lehner type automorphism depending on a fixed primitive pth root of unity as in Sect. 5 or [Gr, Sect. 6]. As w induces an isomorphism between these latter two cohomology groups, we deduce that $\dim_K P^{\mathrm{new}}_k = 2\dim_K H^{\mathrm{new}}_{\mathrm{par}}(k,p)$. This and (8.5) completes the proof. □

We can prove an analogue of Lemma 6.3 in the same way and deduce

Lemma 8.5 *The natural map from* $(M_{k+2,\mathrm{cl}}(p))_\alpha$ *to* $H(k,p)_\alpha$ *is an injection if* $\alpha < k+1$.

This implies

$$\dim_K(S^0_{k+2,\mathrm{cl}}(p))_\alpha \leqq \dim_K H_{\mathrm{par}}(k,p)_\alpha . \tag{8.6}$$

Lemma 8.6 *We have*

$$\dim_K(S^0_{k+2,\mathrm{cl}}(p))_\alpha = \dim_K(S^0_{k+2,\mathrm{cl}}(p))_{k+1-\alpha} . \tag{8.7}$$

Proof. After Corollary 6.4.1, we only have to prove

$$\dim_K(S^{\mathrm{new}}_{k+2,\mathrm{cl}}(p))_\alpha = \dim_K(S^{\mathrm{new}}_{k+2,\mathrm{cl}}(p))_{k+1-\alpha} .$$

But the map $F \mapsto F|w$ is an isomorphism from $(S^{\mathrm{new}}_{k+2,\mathrm{cl}}(p))_\alpha$ onto $(S^{\mathrm{new}}_{k+2,\mathrm{cl}}(p))_{k+1-\alpha}$ generalizing [Gr, Prop. 6.14]. □

Lemma 8.7 *The map from $(S^0_{k+2,\mathrm{cl}}(p))_\alpha$ to $H_{\mathrm{par}}(k,p)_\alpha$ is an isomorphism if $\alpha < k+1$.*

Proof. After Lemma 6.5, it suffices to prove that the map from $(S^{\mathrm{new}}_{k+2,\mathrm{cl}}(p))_\alpha$ to $H^{\mathrm{new}}_{\mathrm{par}}(k,p)_\alpha$ is an isomorphism if $\alpha < k+1$. But this follows from (8.1)–(8.4), (8.6) and (8.7). □

Theorem 8.1 will now follow from the following proposition as Theorem 6.1 did from Proposition 6.6.

Proposition 8.8 *The map from $(M_{k+2,\mathrm{cl}}(p))_\alpha$ to $H(k,p)_\alpha$ is an isomorphism if $\alpha < k+1$.*

Proof. First we observe

$$M_{k+2,\mathrm{cl}}(p) = M_{k+2,\mathrm{cl}} \oplus M^{\mathrm{new}}_{k+2,\mathrm{cl}}(p) ,$$
$$M^{\mathrm{new}}_{k+2,\mathrm{cl}}(p) = E^{\mathrm{new}}_{k+2,\mathrm{cl}}(p) \oplus S^{\mathrm{new}}_{k+2,\mathrm{cl}}(p)$$

and

$$E^{\mathrm{new}}_{k+2,\mathrm{cl}}(p) = (E^{\mathrm{new}}_{k+2,\mathrm{cl}}(p))_0 \oplus (E^{\mathrm{new}}_{k+2,\mathrm{cl}}(p))_{k+1} ,$$

where $E_{k+2,\mathrm{cl}}(p)$ is the space of Eisenstein series of weight $k+2$ over K on $X_1(Np)$. Thus, after Proposition 6.6, all we need to prove is that the natural map from $(E^{\mathrm{new}}_{k+2,\mathrm{cl}}(p))_0$ to $H^{\mathrm{new}}(k,p)/H^{\mathrm{new}}_{\mathrm{par}}(k,p)$ is an isomorphism. We can show that this map is an injection using Lemma 8.5. Hence we must show

$$\dim_K H^{\mathrm{new}}(k,p)/H^{\mathrm{new}}_{par}(k,p) = \dim_K(E^{\mathrm{new}}_{k+2,\mathrm{cl}}(p))_0 = (p-2)C . \tag{8.8}$$

But now we observe that the map from $H^{\mathrm{new}}(k,p)/H^{\mathrm{new}}_{\mathrm{par}}(k,p)$ to

$$H(k,p)/(H(k) + H_{\mathrm{par}}(k,p)) \tag{8.9}$$

is an isomorphism. Let $H(k,p)_{]SS(p)[}$ denote the kernel of the natural map from $H(k,p)$ to $H(k,p)(]SS(p)[)$. Then as $H(k) + H(k,p)_{]SS(p)[} = H(k,p)$, since $H(k)$ surjects onto $H(k)(S) \cong H(k,p)(]SS(p)[)$, we see that the group in (8.9) is isomorphic to $H(k,p)_{]SS(p)[}/(H(k)_{]SS[} + H_{\mathrm{par}}(k,p))$. Thus we have a short exact sequence

$$0 \to H(k)_{]SS(p)[}/H_{\mathrm{par}}(k) \to H(k,p)_{]SS(p)[}/H_{\mathrm{par}}(k,p) \\ \to H^{\mathrm{new}}(k,p)/H^{\mathrm{new}}_{\mathrm{par}}(k,p) \to 0\,.$$

On the one hand, we know by the proof of Lemma 5.3, that

$$\dim_K H(k)_{]SS(p)[}/H_{\mathrm{par}}(k) = \begin{cases} C-1 & \text{if } k=0\,, \\ C & \text{otherwise}\,. \end{cases} \tag{8.10}$$

On the other hand, $H(k,p)_{]SS(p)[}/H_{\mathrm{par}}(k,p)$ is isomorphic to the image of $\mathbf{H}^1(Y, \Omega^{\bullet}(\mathscr{G}_k) \otimes \mathscr{I})$ in $\mathbf{H}^1(Y, \Omega^{\bullet}(\mathscr{G}_k))$ where $\mathscr{I}$ is the ideal sheaf on Y of $C(p) \cap W_1(p)$. Analyzing the relevant long exact sequence as we did in the proof of Lemma 5.3 and using the fact that $\#(C(p) \cap W_1(p)) = (p-1)C$ we see that

$$\dim_K(H(k,p)_{]SS(p)[}/H_{\mathrm{par}}(k,p)) = \begin{cases} (p-1)C-1 & \text{if } k=0\,, \\ (p-1)C & \text{otherwise}\,. \end{cases}$$

This combined with (8.10) establishes (8.8) which concludes the proof. □

In [C-HL], we generalize these results to modular forms on $X_1(Np^n)$ for $n \geqq 1 \in \mathbf{Z}$.

9. Convergent forms of level *Np* and Serre modular forms

Let $t : (\mathbf{Z}/p\mathbf{Z})^* \to \mu_{p-1}(\mathbf{Q}_p)$ denote the Teichmüller character.

Theorem 9.1 *The space of convergent forms of level $\Gamma_1(Np)$, weight k and eigen-character t^i at p is naturally isomorphic to the space of Serre modular forms of level $\Gamma_1(N)$ and weight $(k, k+i) \in \mathbf{Z}_p \times \mathbf{Z}/(p-1)\mathbf{Z}$.*

Compare Theorem 12 of [S] which is a corollary of this.

Lemma 9.2 *There is a unique overconvergent modular form D_p of level $\Gamma_1(Np)$ such that for $d \in (\mathbf{Z}/p\mathbf{Z})^*$*

$$D_p(q) = (E_{p-1}(q))^{1/(p-1)}\,,$$
$$D_p|\langle d\rangle_p = t(d)^{-1} D_p\,.$$

Proof. Uniqueness follows from the fact that a form is determined by its q-expansion. There is a weight one form a on the Igusa curve $I_1(N)$ such that $a^{p-1} = \bar{E}_{p-1}$, $a(q) = 1$ and $a|\langle d\rangle = d^{-1}a$ by [Gr, Prop. 5.2]. Let $\tilde{a}$ be a lifting of a to $M_1(\Gamma_1(Np))$ over $\mathbf{Z}_p[\mu_p(\overline{\mathbf{Q}}_p)]$. By replacing $\tilde{a}$ with $(1/(p-1))\sum_d t(d)\tilde{a}\langle d\rangle$ we may suppose $\tilde{a}|\langle d\rangle = t(d)^{-1}a$. It follows, in particular, that $H =: \tilde{a}^{p-1}/E_{p-1}$ is an element of $M_0(\Gamma_1(N))$ and $H(q) \equiv 1 \bmod \pi$ on Z. Hence there exists an overconvergent function h of level $\Gamma_1(N)$ such that $h^{p-1} = H$. Finally, it is clear we may take D_p to be $\tilde{a}/h$. □

We note that D_p is independent of N in an obvious sense. Does D_p extend to an element of $M_1(\Gamma_1(Np))$?

Proof of Theorem. As $D_p(q)$ is the limit of the sequence

$$\{E_{p-1}(q)^{(p^n-(p^n-1)/(p-1))}\}$$

as n tends to infinity, $D_p(q)$ is a Serre modular form of weight (1,0). Suppose $F \in \omega^k(Z_1(Np))$ and has eigencharacter t^i. Then $FD_p^i \in \omega^{k+i}(Z)$. By [K-pMF, Thm. 4.5.1], this space is naturally isomorphic to the space of Serre modular forms of weight $(k+i, k+i)$. Hence $F(q)$ is a Serre modular form, of weight $(k+i, k+i) - (i, 0) = (k, k+i)$. Since this procedure is obviously invertible we obtain the theorem. $\square$

We could also have proven this result using the classical form labelled $E^*_{(1,0)}$ in [S] in place of D_p.

Acknowledgements. We thank Barry Mazur for encouraging us to think about this problem and Fernando Gouvêa for helpful conversations.

References

[A] Adolphson, A.: The U_p operator of Atkin on modular forms of level three. III. J. Math. **24**, 49–60 (1980)

[BGR] Bosh, S., Güntzer, U., Remmert, R.: Non-Archimedian Analysis. Berlin Heidelberg New York: Springer 1984

[B1] Baldassarri, F.: Comparison entre la cohomologie algébraic et la cohomology p-adique rigide à coefficients dans un module différential I. Invent. Math. **87**, 83–99 (1987)

[B2] Baldassarri, F.: Comparison entre la cohomologie algébraic et la cohomology p-adique rigide à coefficients dans un module différential II. Math. Ann. 417–439 (1988)

[BC] Baldassarri, F., Chiarellotto, B.: Algebraic versus rigid cohomology with logarithmic coefficients, to appear in the proceedings of the Barsotti conference

[RLC] Coleman, R.: Reciprocity laws on curves. Composition **72**, 205–235 (1989)

[pSI] Coleman, R.: A p-adic Shimura isomorphism and p-adic periods of modular forms. Contemp. Math. **165**, 21–51 (1994)

[IP] Coleman, R.: A p-adic inner product on elliptic modular forms, to appear in the proceedings of the Barsotti conference

[C-BFM] Coleman, R.: P-adic Banach spaces and families of Modular forms, to appear.

[C-HL] Coleman, R.: Classical and overconvergent forms of higher level, to appear in Journée Arithmetique

[CGJ] Coleman, R., Gouvêa, F., Jochnowitz, N.: E_2, Θ and U. IMRN **1**, 23–41 (1995)

[D] Dwork, B.: The U_p operator of Atkin on modular forms of level 2 with growth conditions. SLN **350**, 57–67 (1972)

[I] Ihara, Y.: Hecke polynomials as congruence ζ functions in the elliptic modular case. Ann. Math. 267–295 (1967)

[FJ] Faltings, G., Jordan, B.: Crystalline cohomology and $GL(2, \mathbf{Q})$ (to appear)

[G-ApM] Gouvêa, F.: Arithmetic of p-adic modular forms. SLN **1304**, (1988)

[G] Gouvêa, F.: Continuity properties of modular forms, elliptic curves and related topics. CRM Proceedings and Lecture Notes, AMS **4**, 85–99 (1994)

[GM-F] Gouvêa, F., Mazur, B.: Families of modular eigenforms. Math. Comput. **58**, 793–805 (1992)

[GM-CP] Gouvêa, F.: On the characteristic power series of the U operator. Ann. del'institut Fourier **43**, 301–312 (1993)

[Gr] Gross, B.: A tameness criterion for galois representations attached to modular forms (mod p). Duke Math. J. **61**, 445–516 (1990)

[H] Hida, H.: Iwasawa modules attached to congruences of cusp forms. Ann. Sci. Ec. Norm. Sup. **19**, 231–273 (1986)

[K-pMF] Katz, N.: p-adic properties of modular forms, Modular functions of one variable III. SLN **350**, 69–190 (1972)

[K-TD] Katz, N.: Travaux de Dwork, Seminaire Bourbaki, 1971/72, 167–200 (1973)

[K-RT] Katz, N.: The regularity theorem in algebraic geometry, Actes. Congrès Intern. Math. **1**, 437–443 (1970)

[Ki] Kiehl, R.: Theorem A und Theorem B in der nichtarchimedischen Funktiontheorie. Invent. Math. **2**, 256–273 (1967)

[Ko] Koike, M.: On p-adic properties of the Eichler–Selberg trace formula II. Nagoya Math. J. **64**, 87–96 (1976)

[M] Miyake, T.: Modular Forms. Berlin Heidelberg New York: Springer 1989

[R] Ribet, K.: Galois representations attached to eigenforms with nebentypus, Modular functions of one variable V. SLN **601**, 17–36 (1976)

[Sch] Schneider, P.: Points of rigid analytic varieties. J. Reine Angew. Math. 434, 127–157 (1993)

[S-B] Serre, J-P.: Endomorphismes complètements continues des espaces de Banach p-adiques. Publ. Math. IHES **12**, 69–85 (1962)

[S] Serre, J-P.: Formes modulaires et fonctiones zeta p-adiques, Modular functions of one variable III. SLN **350**, 191–268 (1972)

[Sh] Shimura, G.: Sur les integrals attachées aux formes automorphes. J. Math. Soc. Japan **11**, 291–311 (1959)

[Sh-CM] Shimura, G.: On elliptic curves with complex multiplication as factors of the Jacobians of modular function fields. Nagoya Math. J. **43**, 199–208 (1971)