



Slope classicality in higher Coleman theory via highest weight vectors in completed cohomology

Sean Howe^{a,1}

Edited by Kenneth Ribet, University of California, Berkeley, CA; received May 12, 2022; accepted August 2, 2022

We give a proof of the slope classicality theorem in classical and higher Coleman theory for modular curves of arbitrary level using the completed cohomology classes attached to overconvergent modular forms. The latter give an embedding of the quotient of overconvergent modular forms by classical modular forms, which is the obstruction space for classicality in either cohomological degree, into a unitary representation of $\mathrm{GL}_2(\mathbb{Q}_p)$. The U_p operator becomes a double coset, and unitarity yields slope vanishing.

p-adic and overconvergent modular forms | completed cohomology | higher Coleman theory | classicality and control theorems

Fix a sufficiently small compact open subgroup $K^p \leq \mathrm{GL}_2(\mathbb{A}_f^{(p)})$ and let $\mathbb{C}_p$ be the completion of an algebraic closure of $\mathbb{Q}_p$. Let $X_1(p^n)/\mathbb{C}_p$ be the smooth compactification of the modular curve parameterizing elliptic curves with a point of exact order p^n and level K^p structure. Everywhere below, we view $X_1(p^n)$ as an adic space over $\mathbb{C}_p$. The closed canonical ordinary locus $X_1(p^n)_e$ is the topological closure of the locus of rank one points parameterizing elliptic curves of ordinary reduction equipped with a point generating the canonical subgroup of level p^n . We write $X_1(p^n)_w = X_1(p^n) \setminus X_1(p^n)_e$ for its open complement (the subscripts e and w refer to the trivial and nontrivial elements of the Weyl group for GL_2).

Writing ω for the modular sheaf, the space $H^0(X_1(p^n)_e, \omega^k)$ is naturally identified with the direct sum of spaces of overconvergent modular forms of weights κ such that $\kappa = z^k \chi$ where χ is a character of $(\mathbb{Z}/p^n\mathbb{Z})^\times$. From the perspective of the higher Coleman theory of Boxer and Pilloni (1, 2), it is natural to also consider the compactly supported cohomology $H_c^1(X_1(p^n)_w, \omega^k)$. These groups are related by the exact sequence of compactly supported topological sheaf cohomology

$$\begin{array}{ccc} 0 \rightarrow H^0(X_1(p^n), \omega^k) & \longrightarrow & H^0(X_1(p^n)_e, \omega^k) \\ & \nwarrow & \\ H_c^1(X_1(p^n)_w, \omega^k) & \longrightarrow & H^1(X_1(p^n), \omega^k) \rightarrow 0 \end{array} \quad [1]$$

arising from the following exact sequence of sheaves on the topological space $X_1(p^n)$ obtained via push-pull of ω^k along the inclusions $j: X_1(p^n)_w \hookrightarrow X_1(p^n)$ and $i: X_1(p^n)_e \hookrightarrow X_1(p^n)$:

$$0 \rightarrow j_* j^{-1} \omega^k \rightarrow \omega^k \rightarrow i_* i^{-1} \omega^k \rightarrow 0.$$

As in refs. 1 and 2 (see also section 2.2 below), there is an operator U_p on each of these spaces induced by a cohomological correspondence and extending a classical double-coset Hecke operator U_p on $H^\bullet(X_1(p^n), \omega^k)$ (up to matching choices of the normalization). For any $s \in \mathbb{R}$ and $\mathbb{C}_p$ vector space V equipped with an action of a linear operator U_p , we can pass to the part $V^{<s}$ of slope less than s , defined to be the span of all generalized eigenspaces of U_p for eigenvalues λ with $|\lambda| > p^{-s}$.

Theorem 1. For $t \in \mathbb{Z} \setminus \{0\}$, Eq. 1 induces isomorphisms

$$H^0(X_1(p^n), \omega^{1+t})^{<|t|} = H^0(X_1(p^n)_e, \omega^{1+t})^{<|t|} \text{ and } H_c^1(X_1(p^n)_w, \omega^{1+t})^{<|t|} = H^1(X_1(p^n), \omega^{1+t})^{<|t|}.$$

In cohomological degree zero, this is a result of Coleman (3, 4). In degree one, this is a result of Boxer and Pilloni (1, 2) (who also reprove Coleman's result).^{*} We give a short proof using the connection between overconvergent modular forms and the completed cohomology of modular curves established in refs. 5 and 6. This provides a perspective on a fundamental result in the *p*-adic theory of automorphic forms: We recall that Coleman's proof (in the degree zero case) is based on an analysis of the de Rham cohomology of modular curves and a clever dimension counting, while the proof of Boxer and Pilloni is based on slope estimates established via an analysis of cohomological correspondences and integral structures on coherent cohomology. Our proof, by contrast, proceeds by embedding the defect to classicality in completed cohomology so that the necessary slope estimates are a trivial consequence of unitarity, itself a trivial consequence of the construction of completed cohomology from integral singular cohomology (see Remark 1 for the origins of this approach in Emerton's classicality for Jacquet modules). This depends on strong nondegeneracy results of refs. 5 or 6, but the actual construction of the cohomology classes is completely explicit, so that our proof of Theorem 1 reduces to elementary matrix computations.

1. Proof of Theorem 1

Theorem 1 is an immediate consequence of Lemma 1 below, itself an immediate consequence of the results of refs. 5 or 6.

Let X be the infinite-level (compactified) modular curve of prime-to- p -level K^p . It admits an action of $\mathrm{GL}_2(\mathbb{Q}_p)$ and, by Scholze's primitive comparison (see ref. 6, corollary 4.4.3), $H^1(X, \mathcal{O}_X)$ is identified with the $\mathbb{C}_p$ -completed cohomology of the tower of modular curves of prime-to- p -level K^p . We need only that it is a Banach space with a unitary action of $\mathrm{GL}_2(\mathbb{Q}_p)$, which follows because the unit ball, that is, the image of $H^1(X, \mathcal{O}_X^+)$, is preserved by $\mathrm{GL}_2(\mathbb{Q}_p)$.

Author affiliations: ^aDepartment of Mathematics, University of Utah, Salt Lake City, UT 84112

Author contributions: S.H. designed research, performed research, and wrote the paper.

The author declares no competing interest.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

¹Email: sean.howe@utah.edu.

Published November 2, 2022.

^{*}In ref. 1 and the introduction of ref. 2, this result is stated at level $\Gamma_0(p)$ using the smaller group of cohomology with support in $X_0(p)_W^{\mathrm{ord}}$. It is immediate from the results of *loco citato* that this smaller space has the same finite slope part, and arbitrary level is treated in ref. 2, theorem 5.12.2.

Lemma 1. Let U_p denote the Hecke operator of refs. 1 and 2 (the normalization depends on the weight; see section 2), and let $N \leq \mathrm{GL}_2$ be the group of upper triangular unipotent matrices. For $t \neq 0$, the cup products of ref. 5 give an embedding

$$H^0(X_1(p^n)_e, \omega^{1+t})/H^0(X_1(p^n), \omega^{1+t}) \hookrightarrow H^1(X, \mathcal{O}_X)^{N(\mathbb{Z}_p)}$$

matching U_p with the double coset $p^{|t|} \cdot [N(\mathbb{Z}_p) \mathrm{diag}(p, 1) N(\mathbb{Z}_p)]$.

Proof of Theorem 1, assuming Lemma 1: Consider

$$p^{|t|} \cdot [N(\mathbb{Z}_p) \mathrm{diag}(p, 1) N(\mathbb{Z}_p)] = p^{|t|} \sum_{i=0}^{p-1} \begin{bmatrix} p & i \\ 0 & 1 \end{bmatrix}$$

acting on $H^1(X, \mathcal{O}_X)^{N(\mathbb{Z}_p)}$. Because the action of $\mathrm{GL}_2(\mathbb{Q}_p)$ is unitary, it has operator norm $\leq 1/p^{|t|}$, so the slope $< |t|$ part vanishes in $H^1(X, \mathcal{O}_X)^{N(\mathbb{Z}_p)}$. If we write

$$Q_t := H^0(X_1(p^n)_e, \omega^{1+t})/H^0(X_1(p^n), \omega^{1+t}),$$

then, combining the above with Lemma 1, we find that, for $t \neq 0$, $Q_t^{<|t|} = 0$. To obtain Theorem 1, we split Eq. 1 for $k = 1 + t$ into two short exact sequences,

$$\begin{aligned} 0 \rightarrow H^0(X_1(p^n), \omega^{1+t}) \rightarrow H^0(X_1(p^n)_e, \omega^{1+t}) \rightarrow Q_t \rightarrow 0 \text{ and} \\ 0 \rightarrow Q_t \rightarrow H_c^1(X_1(p^n)_w, \omega^{1+t}) \rightarrow H^1(X_1(p^n), \omega^{1+t}) \rightarrow 0. \end{aligned}$$

Taking the slope $< |t|$ part yields the isomorphisms—for the first sequence this is immediate, since this functor is always left exact, and, for the second sequence, right exactness follows from compactness of the U_p operator on overconvergent forms. $\square$

Remark 1. As recalled in Remark 3 below, for $t > 0$, the embedding of Lemma 1 arises from a highest weight vector in the irreducible submodule of a Verma module with algebraic quotient. The argument is then essentially the same as Emerton's proof of the classicality theorem for locally analytic Jacquet modules (ref. 7, theorem 4.4.5).

It thus remains only to prove Lemma 1. This is essentially immediate from the results of refs. 5 or 6, once the $\mathrm{GL}_2(\mathbb{Q}_p)$ actions are matched up. This matching is actually a bit subtle, as there are multiple possible conventions for the Hodge–Tate period map and the equivariant structure on the modular sheaf. Any set of choices gives the same $\mathrm{GL}_2(\mathbb{Q}_p)$ -action modulo inverse transpose and some determinants, so, often, the precise choices are irrelevant. Here, however, we must follow a power of p coming from the action of $\mathrm{diag}(p, 1)$, so it is crucial to screw our heads on exactly right on this point. In the next section, we fix normalizations, then prove Lemma 1.

2. Normalizations and Proof of Lemma 1

2.1. Choices. We fix the action of $\mathrm{GL}_2(\mathbb{Q}_p)$ on X so that, over the noncompactified infinite-level curve Y , $\mathrm{GL}_2(\mathbb{Q}_p) = \mathrm{Aut}(\mathbb{Q}_p^2)$ acts by composition with the trivialization of the Tate module of the universal elliptic curve; that is, we use the action on the homological normalization of the moduli problem. This differs by an inverse transpose from the cohomological normalization, where the action is on the trivialization of the first étale cohomology of the universal elliptic curve.

We take the Hodge–Tate period map $\pi_{\mathrm{HT}} : X \rightarrow \mathbb{P}^1$ so that $\pi_{\mathrm{HT}}|_Y$ is the classifying map for the Hodge–Tate line inside of

the first étale cohomology of the universal elliptic curve. Thus, over Y , we have a $\mathrm{GL}_2(\mathbb{Q}_p)$ -equivariant commuting diagram,

$$\begin{array}{ccccc} \mathcal{O}_X e_1 \oplus \mathcal{O}_X e_2 & \xrightarrow{\sim} & \mathcal{O}_X \otimes V_p(E) & \longrightarrow & \omega_{E^\vee} \\ \pi_{\mathrm{HT}}^* \uparrow & & \pi_{\mathrm{HT}}^* \uparrow & & \pi_{\mathrm{HT}}^* \uparrow \\ \mathcal{O}_{\mathbb{P}^1} x \oplus \mathcal{O}_{\mathbb{P}^1} y & \xrightarrow{\sim} & \mathcal{O}_{\mathbb{P}^1} \otimes H^0(\mathbb{P}^1, \mathcal{O}_{\mathbb{P}^1}(1)) & \longrightarrow & \mathcal{O}_{\mathbb{P}^1}(1). \end{array}$$

Here e_1 and e_2 are the universal basis for the Tate module $V_p(E) = T_p(E)[1/p]$, x and y are the standard basis for $H^0(\mathbb{P}^1, \mathcal{O}_{\mathbb{P}^1}(1))$ so that homogeneous coordinates are $[x : y]$, and $E^\vee$ denotes the dual of the universal elliptic curve. Of course, there is a canonical isomorphism $E^\vee \cong E$ inducing $\omega_{E^\vee} \cong \omega_E$; however, this isomorphism does not respect the natural $\mathrm{GL}_2(\mathbb{Q}_p)$ -equivariant structures! Equivariantly,

$$\omega_{E^\vee} = \omega_E \otimes |\det|^{-1},$$

where, here, $|\det|$ comes from the action of isogenies on $H^1(E, \Omega_E)$. Note that this twist is actually on the entire $\mathrm{GL}_2(\mathbb{A}_f)$ action (via $(g_\ell)_\ell \mapsto \prod_\ell |\det(g_\ell)|_\ell^{-1}$), so that the distinction between these equivariant structures also determines the normalization of the prime-to- p Hecke operators. Below we will continue, as in the introduction, to write simply ω for the modular sheaf, with the understanding that we have adopted the equivariant structure described above.

Under the natural map to $X \rightarrow X_1(p^n)$, $X_1(p^n)_e$ is the image of $\pi_{\mathrm{HT}}^{-1}([0 : 1])$, and the action of $\mathrm{GL}_2(\mathbb{Q}_p)$ on $\langle x, y \rangle = H^0(\mathbb{P}^1, \mathcal{O}_{\mathbb{P}^1}(1))$ is by the standard representation

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot x = ax + cy \text{ and } \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot y = bx + dy.$$

2.2. The U_p^{naive} Operator. The operator U_p^{naive} at level $X_1(p^n)$ of refs. 1 and 2 is defined using the correspondence C parameterizing degree p isogenies $\psi : (E_1, P_1) \rightarrow (E_2, P_2)$ (here we suppress prime-to- p -level structure from the notation). Writing the two obvious projections as $p_1, p_2 : C \rightarrow X_1(p^n)$, U_p^{naive} is defined on ω^k as $\mathrm{tr} \circ p_{1!} \circ \psi^* \circ p_2^*$. Given a geometric point (E, P) that is not a cusp and a nonzero differential η on E , we can compute $(U_p^{\mathrm{naive}} f)(E, P, \eta)$ as follows: First, choose a basis (e_1, e_2) of $T_p(E)$ such that e_1 reduces to $P \bmod p^n$. Then, for $0 \leq i \leq p-1$, write

$$\psi_i : E \rightarrow E_i := E / \langle i\bar{e}_1 + \bar{e}_2 \rangle,$$

where $\bar{e}_i$ denotes the image of e_i in $E[p]$. Then $\psi_i^* : \omega_{E_i} \rightarrow \omega_E$ is invertible, and

$$(U_p^{\mathrm{naive}} f)(E, P, \eta) = \sum_{i=0}^{p-1} f(E_i, \psi_i(P), (\psi_i^*)^{-1} \eta). \quad [2]$$

We will now realize this same U_p^{naive} as a double-coset operator: Let B denote the upper triangular Borel in GL_2 . The space of overconvergent modular forms of weight k at any finite level $\Gamma_1(p^n)$ is naturally embedded as the $B_1(p^n) = \Gamma_1(p^n) \cap B(\mathbb{Q}_p)$ invariants in the $B(\mathbb{Q}_p)$ representation

$$M_k^\dagger := H^0([0 : 1], (\pi_{\mathrm{HT}*} \pi_{\mathrm{HT}}^* \mathcal{O}(k))^{\mathrm{sm}}),$$

where, here, the superscript sm denotes the subsheaf of $\pi_{\mathrm{HT}*} \pi_{\mathrm{HT}}^* \mathcal{O}(k) = \pi_{\mathrm{HT}*} \omega^k$ whose sections over a quasi-compact open V are those with locally constant orbit map for the action of the stabilizer of V in $\mathrm{GL}_2(\mathbb{Q}_p)$ (i.e., sections fixed by some compact open subgroup of $\mathrm{GL}_2(\mathbb{Q}_p)$, i.e., sections coming from finite level). For more on this construction, see ref. 5, section 3.1.

The space $M_k^\dagger$ contains the space of classical modular forms

$$M_k^{\text{cl}} = H^0(\mathbb{P}^1, (\pi_{\text{HT}*} \pi_{\text{HT}}^* \mathcal{O}(k))^{\text{sm}})$$

$B(\mathbb{Q}_p)$ equivariantly by restriction. The action of

$$U := N(\mathbb{Z}_p) \text{diag}(p, 1) N(\mathbb{Z}_p) = \sum_{i=0}^{p-1} \begin{bmatrix} p & i \\ 0 & 1 \end{bmatrix} =: \sum_{i=0}^{p-1} A_i$$

on $M_k^{\dagger, N(\mathbb{Z}_p)}$ is identified with U_p^{naive} on $M_k^{\dagger, B_1(p^n)} = H^0(X_1(p^n), \omega^k)$ —we explain this computation now. It suffices to check at geometric points away from cusps, so we can compare with the explicit formula of Eq. 2 for U_p^{naive} . Now,

$$\begin{aligned} U(E, e_1, e_2, \eta) &= \sum_{i=0}^{p-1} (A_i \cdot f)(E, e_1, e_2, \eta) \\ &= \sum_{i=0}^{p-1} f(E, pe_1, ie_1 + e_2, \eta), \end{aligned}$$

and, by the commuting diagram

$$\begin{array}{ccccc} \mathbb{Q}_p^2 & \xrightarrow{A_i} & \mathbb{Q}_p^2 & \xrightarrow{e_1, e_2} & V_p E \\ & \searrow \psi_i(e_1), \psi_i(\frac{1}{p}(ie_1 + e_2)) & & \uparrow \psi_i^\vee & \\ & & & V_p(E_i), & \end{array}$$

we see that $f(E, pe_1, ie_1 + e_2, \eta)$ is equal to $f(E_i, \psi_i(e_1), \psi_i(\frac{1}{p}(ie_1 + e_2)), \eta_i)$ for some η_i . One might guess that η_i is $(\psi_i^\vee)^* \eta$, but it is not! This would hold if we used the ω_E -equivariant structure, but, in the $\omega_{E^\vee}$ -equivariant structure, we must replace $\psi_i^\vee$ with its dual ψ_i so that $\eta_i = (\psi_i^*)^{-1} \eta$. Thus we recover the right-hand side of Eq. 2.

2.3. Proof of Lemma 1 and Concluding Remarks. Recall (ref. 1, sentence preceding theorem 5.13) the normalized operator

$$U_p := \begin{cases} p^{-1} U_p^{\text{naive}} & \text{if } k \geq 1 \\ p^{-k} U_p^{\text{naive}} & \text{if } k \leq 1 \end{cases} \text{ so } U_p^{\text{naive}} = \begin{cases} p U_p & \text{if } k \geq 1 \\ p^k U_p & \text{if } k \leq 1. \end{cases}$$

appearing in the statement of Lemma 1 (see also Remark 4).

Proof of Lemma 1. We first treat the case $k = 1 + t \geq 2$. Then, for any $s \in M_k^\dagger$, s/x^k is a section of $\mathcal{O}_X$ defined on the preimage of a punctured neighborhood of $[0 : 1]$ under π_{HT} . It thus determines a Čech cohomology class $[s/x^k]$ in $H^1(X, \mathcal{O}_X)$. Then, ref. 5, theorem A implies that $s \mapsto [s/x^k]$ induces an injection $M_k^\dagger / M_k^{\text{cl}} \hookrightarrow H^1(X, \mathcal{O}_X)$; actually, in ref. 5, the results are stated using cusp forms and compactly supported completed cohomology, but, given the identification of $H^1(X, \mathcal{O}_X)$ with completed cohomology, one obtains the desired statement by the same arguments. The map is $B(\mathbb{Q}_p)$ equivariant if one twists the action on $M_k^\dagger$ by

$$\begin{bmatrix} a & b \\ 0 & c \end{bmatrix} \mapsto a^{-k}$$

(the twist comes from the action on x^k , of course!). We deduce that $p^{-k} U_p^{\text{naive}} = p^{-k} (p U_p) = p^{-t} U_p$ is identified with $[N(\mathbb{Z}_p) \text{diag}(p, 1) N(\mathbb{Z}_p)]$, as desired.

We now treat the case $k = 1 + t \leq 0$. In this case, ref. 5, theorem A shows that $s \mapsto [s/(xy^t)]$ induces an embedding $M_k^\dagger / M_k^{\text{cl}} \hookrightarrow H^1(X, \mathcal{O}_X)$. Actually, here one must be slightly more careful invoking the arguments of ref. 5, which are stated with cusp forms, in the case $k = 0$: Of course, $M_k^{\text{cl}} = 0$ when $k < 0$, but, when $k = 0$, we have that M_0^{cl} is the locally constant functions, whereas the cusp forms are still trivial. However, it is elementary to see that M_0^{cl} is in the kernel (as $s/(xy^{-1}) = s/z$ extends to a function on the complement of $[0 : 1]$ where $1/z$ is a local coordinate), and the argument of *loco citato* still establishes an injection on the quotient by M_0^{cl} . The embedding is $B(\mathbb{Q}_p)$ equivariant if we twist $M_k^\dagger$ by

$$\begin{bmatrix} a & b \\ 0 & c \end{bmatrix} \mapsto a^{-1} c^{-t},$$

where, again, the twist comes from the action on xy^t . We deduce that $p^{-1} U_p^{\text{naive}} = p^{-1} (p^k U_p) = p^t U_p$ is matched with $[N(\mathbb{Z}_p) \text{diag}(p, 1) N(\mathbb{Z}_p)]$, concluding the proof. $\square$

Remark 2. Lemma 1 can also be deduced from ref. 6, theorem 1.0.1, and this has the advantage that it is stated already with completed cohomology instead of compactly supported completed cohomology and cusp forms. We have used ref. 5 above because it was easier to check carefully our own normalizations!

Remark 3. The vectors used for $k \leq 0$ also exist for $k \geq 2$, where they induce an injection on all overconvergent modular forms. The same argument then recovers the fact that U_p has nonnegative slopes when $k \geq 2$ (of course, it is much simpler to deduce this from the action on q expansions!). Representation theoretically, this vector comes from a highest weight Verma module, which, when $k \geq 2$, admits an algebraic quotient; the classical forms are exactly those that factor through the algebraic quotient, and the vector we used above for the $k \geq 2$ case is the lower highest weight vector generating the kernel. This perspective is explained in ref. 5.

Remark 4. The reason one is led to use different normalizations depending on $k \leq 0$ or $k \geq 2$ is mostly explained by the form of the Hodge–Tate sequence

$$\text{Lie} E(1) \rightarrow T_p(E) \otimes \mathcal{O}_Y \rightarrow \omega_{E^\vee}.$$

Indeed, since we are using the double-coset operator for the p -integral matrix $\text{diag}(p, 1)$ acting on $V_p E$, to obtain an operator with nonnegative slopes, it is natural to use the equivariant structure from $\omega_{E^\vee}$ for $k = 1$ and from $\text{Lie} E = \omega_E^{-1}$ for $k = -1$, and similarly for larger $|k|$ by taking symmetric powers of $T_p(E)$. The equivariant structures differ by an absolute value of the determinant, which manifests here as different powers of p for the double-coset operator coming from $\text{diag}(p, 1)$.

Data, Materials, and Software Availability. There are no data underlying this work.

ACKNOWLEDGMENTS. We thank George Boxer for providing the precise reference, in “Higher Coleman theory,” for the higher-level case of Theorem 1.

1. G. Boxer, V. Pilloni, Higher Hida and Coleman theories on the modular curve. *arXiv [Preprint]* (2020). <https://doi.org/10.48550/arXiv:2002.06845> (Accessed 1 May 2022).
2. G. Boxer, V. Pilloni, Higher Coleman theory. *arXiv [Preprint]* (2020). <https://doi.org/10.48550/arXiv:2110.10251> (Accessed 1 May 2022).
3. R. F. Coleman, Classical and overconvergent modular forms. *Invent. Math.* **124**, 215–241 (1996).
4. R. F. Coleman, Classical and overconvergent modular forms of higher level. *J. Théor. Nombres Bordeaux* **9**, 395–403 (1997).

5. S. Howe, Overconvergent modular forms are highest-weight vectors in the Hodge–Tate weight zero part of completed cohomology. *Forum Math. Sigma* **9**, E17 (2021).
6. L. Pan, On locally analytic vectors of the completed cohomology of modular curves. *Forum Math. Pi* **10**, E7 (2022).
7. M. Emerton, Jacquet modules of locally analytic representations of p -adic reductive groups. I. Construction and first properties. *Ann. Sci. École Norm. Sup.* **39**, 775–839 (2006).