The Eigencurve

R. COLEMAN AND B. MAZUR

in memory of Bernard Dwork

Let p be a prime number and $\mathbf{C}_p$ the completion of an algebraic closure of the field $\mathbf{Q}_p$ of p-adic numbers. Let N be an integer relatively prime to p. To describe our main object, we assume that $p > 2$, that the group of units in the ring $\mathbf{Z}/N\mathbf{Z}$ is of order prime to p, and we restrict our attention (at least in this introduction) to classical modular cuspidal eigenforms $f = \sum_{n=1}^{\infty} a_n q^n$ on $\Gamma_0(pN)$ of weight $k \geq 2$, with Fourier coefficients in $\mathbf{C}_p$, and normalized so that $a_1 = 1$. By "eigenform" let us mean eigenform for the Hecke operators T_ℓ for primes ℓ not dividing pN, for the Atkin-Lehner operators U_q for primes q dividing pN and for the diamond operators $\langle d \rangle$ for integers d prime to Np. By the **slope** of such an eigenform we mean the non-negative rational number $\sigma = \mathrm{ord}_p(a_p)$, where a_p is (both) the p-th Fourier coefficient of f and the U_p-eigenvalue of f. Assume that the newform associated to f is either a newform for $\Gamma_0(N)$ or $\Gamma_0(pN)$ (the latter case can occur only if $\sigma = (k-2)/2$).

By the work of Hida (cf. [H-ET] for an exposition of this theory, and for further bibliography given there) one knows that any such eigenform f of weight $k \geq 2$ and slope 0 is a member of a p-adic analytic family f_κ of (overconvergent) p-adic modular eigenforms of slope 0 parameterized by their p-adic weights κ (and, such that $f_k = f$) for κ ranging through a small p-adic neighborhood of k in (p-adic) weight space. Let us call this result *p-adic analytic variation of slope* 0 *eigenforms.* But Hida's results are, in fact, much more precise: If $\Lambda_N := \mathbf{Z}_p[[(\mathbf{Z}/N\mathbf{Z})^* \times \mathbf{Z}_p^*]]$, Hida constructs a finite flat Λ_N-algebra (let us call it $\mathbf{T}^o_{p,N}$) which is universal, in a certain sense, for slope 0 (overconvergent) eigenforms of tame level N, and such that the associated rigid analytic space to $\mathbf{T}^o_{p,N}$ (let us call it $C^o_{p,N}$) is the rigid-analytic space parameterizing p-adic analytic families of slope 0 eigenforms. If $\mathcal{W}_N$ is weight space, i.e., the rigid analytic space associated to Λ_N, then the p-adic families f_κ alluded to above are obtained from the finite flat projection $C^o_{p,N} \to \mathcal{W}_N$, Hida having proved that this mapping is

Reprinted from 'Galois Representations in Arithmetic Algebraic Geometry', edited by A. J. Scholl & R. L. Taylor.

étale at any classical modular point of weight ≥ 2 in $C^o_{p,N}$.

The natural question arising from this work of Hida for eigenforms of slope 0 is to find the appropriate generalization of that theory valid for arbitrary finite slope eigenforms. The slope 0 theory has a significant simplifying advantage over the general theory in that there is available a clean idempotent operator (call it e^o) projecting to the slope 0 part of the theory. This idempotent e^o, a p-adic idempotent reminding one a bit of the "holomorphic projector" in classical analysis, is a bounded operator on the Banach spaces involved, and in fact it has no denominators and therefore acts as an idempotent on all aspects of the theory (e.g., parabolic cohomology, as well as spaces of modular forms); moreover the image of this idempotent is of finite type over whatever is the natural base ring.

In [C-BMF] a satisfactory analogue of Hida's p-adic analytic variation theorem for slope 0 eigenforms was established for finite slope classical eigenforms (at least for those satisfying a mild condition; cf. Cor. B5.7.1 of [C-BMF]).

The ultimate aim of the theory developed in this article is to provide a more global counterpart to the work done in [C-BMF] and construct a rigid analytic curve $C_{p,N}$ (analogous to $C^o_{p,N}$) which parameterizes *all* finite slope overconvergent p-adic eigenforms of tame level N, and to study its detailed geometry: in particular, we wish to understand the nature of the projection of $C_{p,N}$ to weight space $\mathcal{W}_N$.

For reasons of space, and time, we do this only in the case of $p > 2$, and tame level $N = 1$ in the present article. We do, however, treat noncuspidal eigenforms as well as eigenforms on $\Gamma_1(p^m)$, as opposed to the more restricted class of eigenforms delineated at the beginning of this introduction. For more precise, yet still introductory, statements concerning our main results, the reader might turn to section 1.3 (and in particular to the Theorems A,B,C,D,E,F,G formulated there). For the rest of this introduction, we suppose that $p > 2$, and discuss the case of tame level $N = 1$.

We construct a rigid analytic curve $C_p = C_{p,N=1}$ over $\mathbf{Q}_p$ whose $\mathbf{C}_p$-valued points parameterize *all* finite slope overconvergent p-adic eigenforms of tame level $N = 1$ with Fourier coefficients in $\mathbf{C}_p$. We call C_p *the (p-adic) eigencurve (of tame level $N = 1$)*. Hida's rigid space $C^o_p = C^o_{p,N=1}$ which parameterizes slope 0 eigenforms of tame level 1 occurs as a component part (cf. section 1.2) of our eigencurve C_p, but in contrast to Hida's theory, the natural projection of C_p to weight space is *not* of finite degree. The

eigencurve C_p has a natural embedding

$$c \mapsto (\rho_c, 1/u_c)$$

into the rigid analytic space $X_p \times \mathbf{A}^1$ where X_p is the rigid analytic space attached to the universal deformation ring R_p of certain Galois (pseudo-) representations and $\mathbf{A}^1$ is the affine line. For a discussion of pseudo-representations, see Chapter 5 below. The particular residual pseudo-representations for which R_p is the universal deformation ring we call **p-modular** residual representations (of the Galois group $G_{\mathbf{Q},\{p,\infty\}}$). These are the residual pseudo-representations coming from the classical modular eigenforms of finite slope and level a power of p (See section 5.1 for their definition.); there are only finitely many such residual pseudo-representations, and R_p is a complete semi-local noetherian ring.

If $c \in C_p$ corresponds to the overconvergent eigenform f_c, the first coordinate of c with respect to this embedding is the Galois pseudo-representation $r_c \in X_p(\mathbf{C}_p)$ attached to the eigenform f_c and the second coordinate $(1/u_c)$ is the inverse of $u_c :=$ the U_p-eigenvalue of the eigenform f_c. We define the eigencurve $C_p \subset X_p \times \mathbf{A}^1$ to be the rigid analytic subspace cut out by an ideal generated by certain specific rigid analytic functions, these functions being *Fredholm series over* R_p (cf. 1.2) obtained by pullback to $X_p \times \mathbf{A}^1$ of the characteristic series of certain completely continuous systems of operators (cf. 4.3). In general, a Fredholm series over a complete local noetherian ring R is an entire power series (with constant term 1) in one variable T over R, and we call a *Fredholm variety* over R (see section 1.2 below) a rigid analytic subspace of $X_p \times \mathbf{A}^1$ which is cut out by an ideal generated by a collection of Fredholm series over R (here T is the variable parameterizing the affine line $\mathbf{A}^1$).

Let us refer to the nilreduced rigid analytic space subjacent to the eigencurve as the **reduced eigencurve** C_p^{red}. For a brief discussion of the process of passing to the nilreduction of a rigid analytic space, see section 1.2 below. We define irreducible components there and prove under sufficiently general hypotheses that every point in such a space lies on one.

We show that each irreducible component of the reduced eigencurve C_p^{red} is isomorphic (at least outside of a discrete set of points) to some reduced irreducible Fredholm hypersurface over the Iwasawa ring Λ (i.e., it is a Fredholm variety defined by a single (irreducible) Fredholm series with coefficients in the Iwasawa algebra Λ) via an isomorphism that preserves projection to weight space. Each irreducible component of a Fredholm hypersurface over Λ is again a Fredholm hypersurface over Λ, and it is seen (Cor. 1.3.12 below) to either be isomorphic to the open complement of a

finite set of points in the rigid analytic space attached to a finite flat Λ-algebra, or else to be of infinite degree. When (and if) the first case holds, we would be in a situation very analogous to Hida's theory for slope 0 eigenforms. But so far we *have not yet been able to determine* for any positive slope irreducible component of C_p^{red} which of these two alternatives hold!

It follows from our results that the natural projection of each irreducible component of the reduced eigencurve to weight space is componentwise almost surjective in the sense that its image avoids at most a finite number of $\mathbf{C}_p$-valued points in the component of weight space within which it lands; in particular, each irreducible component of C_p^{red} has eigenforms of all (but a finite number of) weights. For a more precise statement, see Theorem B of 1.5.

It also follows from our results that any convergent eigenform of finite slope and integral weight (of tame level 1) is overconvergent if and only if its q-expansion is approximable p-adically by the q-expansions of classical eigenforms (of tame level 1). (Note that the q-expansions of Serre's p-adic modular forms are the limits of the q-expansions of classical modular (but not necessarily eigen) forms.) For a more precise statement, see Theorem G of 1.5.

In Chapter 7, we construct a second rigid space D, by means of the Banach module theory of [C-BMF], which is evidently a curve. We prove the above assertions (in particular, that it is a curve) about C_p^{red} by proving them about D and showing that

$$D \cong C_p^{\mathrm{red}}.$$

From the its construction, one sees that the projection of D to weight space is *locally in-the-domain finite flat*, meaning that D is covered by admissible affinoid domains $\mathcal{U}$ such that the restriction of projection to weight space to $\mathcal{U}$ is a finite flat mapping of $\mathcal{U}$ onto its image in $\mathcal{W}$.

The image of a rigid analytic morphism can be quite intricate, and, in particular, the projection $C_p^{\mathrm{red}} \to X_p$ has an infinite number of double points (e.g., those coming from the classical modular eigenforms of level 1). The image of C_p^{red} in X_p contains the "infinite ferns" studied in the articles [GM-FM], [GM 3] and [M-IF].

We may pull back the universal pseudo-representation on X_p to C_p^{red}, via the natural projection, to obtain a rigid analytically varying pseudo-representation on C_p^{red} which is realizable, at least on the complement of a certain discrete set of points on C_p^{red} as an $\mathcal{O}_{C_p^{\mathrm{red}}}$-linear continuous Gal $(\bar{\mathbf{Q}}/\mathbf{Q})$-representation unramified outside p with the property that the

restriction of this representation to a $\mathbf{C}_p$-valued point $c \in C_p$ yields the p-adic Galois representation attached to the eigenform f_c.

Open questions. Is C_p reduced? (I.e., does $C_p = D$?) Is C_p smooth? Does C_p^{red} have a finite or an infinite number of components? Do any of these components have infinite genus? Is every component of C_p^{red} of slope > 0 of *infinite* degree over weight space, or are there components of positive slope that are of finite degree over weight space? Do there exist p-adic analytic families of overconvergent eigenforms of finite slope parameterized by a punctured disc, and converging, at the puncture, to an overconvergent eigenform of infinite slope? (If so, one would want to *complete* the eigencurve a bit by including these missing points of infinite slope.) Having constructed the (global) eigencurve allows us to ask more global questions along the lines of conjectures made by one of the authors of the present article, with Fernando Gouvêa, and bears on recent work of Daqing Wan [Wa]: Are the ramification points of the natural projection from C_p to weight space infinite in number? We would like to know where those ramification points are; specifically given a point $c \in C_p$ of weight $\kappa \in \mathcal{W}$ let us say that an affinoid subdomain $\mathcal{U} \subset \mathcal{W}$ containing κ is a *weight-parameter space* for c if there is an affinoid neighborhood $\mathcal{V} \subset C_p$ of c whose natural projection to $\mathcal{U}$ is finite étale. Can one find, given any point c corresponding to a classical eigenform of weight k of slope strictly less than $k-1$, a weight-parameter space for c of radius greater than the inverse of a linear function of the slope of c? One can prove, using Wan's results [Wa], that there are such weight-parameter spaces of radius greater than the inverse of a quadratic function of the slope if the slope is strictly less than $k-1$ and not equal to $(k-1)/2$.

We mentioned above that, excluding a discrete set of points (call this set $\Delta \subset C_p^{\mathrm{red}}$), the reduced eigencurve parametrizes a rigid-analytically varying family of Galois representations, i.e. we have a continuous representation

$$\rho : \mathrm{Gal}\,(\bar{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathcal{O}_{C_p^{\mathrm{red}}-\Delta})$$

Can this family of Galois representations ρ be extended over the excluded set Δ? Is there a continuous representation of Gal $(\bar{\mathbf{Q}}/\mathbf{Q})$ into the group of units in a rigid Azumaya algebra of rank 4 over $\mathcal{O}_{C_p}$ which, in the appropriate sense, extends ρ? Can one construct ρ (or at least ρ restricted to an appropriate $C_p^{\mathrm{red}} - \Delta$) from the cohomology of modular curves in the following sense: Let $\mathcal{H}$ be the polynomial algebra over Λ in the countable set of generators denoted T_ℓ for prime numbers $\ell \neq p$ and an *extra* generator denoted U_p. For a more complete discussion of this Λ-algebra $\mathcal{H}$ see Chapter 6. There is a natural homomorphism of $\mathcal{H}$ to the ring $\mathcal{O}_{C_p}$ of rigid

analytic functions on the eigencurve; in particular, in the discussion below we view $\mathcal{O}_{C_p}$ as $\mathcal{H}$-algebra. Define the discrete abelian group

$$\mathcal{M} := \varinjlim H^1(X_1(p^n); \mathbf{Q}_p/\mathbf{Z}_p),$$

the direct limit of étale cohomology groups of the modular curves $X_1(p^n)$, this limit being compiled via the mappings on cohomology induced from the natural projections $X_1(p^{n+1}) \to X_1(p^n)$ for all $n \geq 1$ (in the present brief description we omit saying *which* natural projections these are). The Galois group $G_{\mathbf{Q},\{p,\infty\}}$ acts on this direct limit in the natural way, as does the algebra $\mathcal{H}$ (each of the generators acting as the corresponding Hecke or Atkin-Lehner operator) and the actions of $\mathcal{H}$ and $G_{\mathbf{Q},\{p,\infty\}}$ commute, allowing us to view $\mathcal{M}$ as an $\mathcal{H}[[G_{\mathbf{Q},\{p,\infty\}}]]$-module. Let $\mathcal{M}^* := \mathrm{Hom}(\mathcal{M}; \mathbf{Q}_p/\mathbf{Z}_p)$ be the pontrjagin dual of $\mathcal{M}$, viewed as compact $\mathcal{H}[[G_{\mathbf{Q},\{p,\infty\}}]]$-module. There are various (possibly equivalent) ways of *spreading the cohomology $\mathcal{H}$-module $\mathcal{M}^*$ over the eigencurve.* The simplest way to do this is to form $V :=$ the completed tensor product of the $\mathcal{H}$-module $\mathcal{M}^*$ with the $\mathcal{H}$-algebra $\mathcal{O}_{C_p}$. Viewing V as quasi-coherent sheaf over the eigencurve, the $G_{\mathbf{Q},\{p,\infty\}}$-action on $\mathcal{M}^*$, which commutes with the action of $\mathcal{H}$, induces a $\mathcal{O}_{C_p}$-linear action on V. Is it the case that, over C_p, or possibly just over some large portion of C_p, the quasi-coherent sheaf V is locally free of rank 2, and the $G_{\mathbf{Q},\{p,\infty\}}$-representation on it is equivalent to the family of representations ρ discussed above? On the converse side it would be good to find an intrinsic operation which cuts out of $\mathcal{M}^*$ precisely those Galois representations which are attached to overconvergent eigenforms. All these questions have analogues when the cohomology group

$$\mathcal{M} := \varinjlim H^1(X_1(p^n); \mathbf{Q}_p/\mathbf{Z}_p)$$

is replaced by (direct limits of) parabolic cohomology of higher weights, and one might ask for a theory, following Hida's work, which deals with all weights. Relevant to this are the results of Hida [H-HA] and Gouvêa [G-ApM].

Does Glenn Stevens' construction of the p-adic L-function of a (p-adic) eigenform on $\Gamma_0(pN)$ (cf. [St]) work well over the eigencurve, and, in particular, does it give an L-function (with the coefficients of its Taylor expansion rigid analytic functions on C_p) which interpolates all the classical p-adic L-functions? We hope to show in a later publication, that at least a weak version of this is true. The locus of zeroes of this L-function, once constructed, would itself be a rigid-analytic curve (call it $\mathcal{L}_p$) admitting a natural rigid analytic projection to C_p. This deserves study. particularly intriguing is the nature of this projection in the neighborhood of a double-zero of the L-function.

It would be interesting to construct a local version of the eigencurve using the theory of crystalline Galois representations, in the following sense. Let X now refer to the rigid-analytic universal deformation space of a fixed absolutely irreducible representation of the Galois group of $\mathbf{Q}_p$ into $GL_2(\mathbf{F}_p)$. It is known that X is a smooth 5-dimensional rigid-analytic ball. By a crystalline point in (the 6-dimensional) rigid-analytic space $X \times \mathbf{A}^1 - \{0\}$, let us mean a pair (x, u) where $x \in X$ classifies a crystalline representation, and $u \in \mathbf{A}^1 - \{0\}$ is one of its the Frobenius eigenvalues in the Fontaine-Dieudonné module attached to the crystalline representation x. (For a general overview of crystalline, semi-stable, and potentially semi-stable representations, see [F].) Is it the case that the rigid-analytic closure of the set of crystalline points in $X \times \mathbf{A}^1 - \{0\}$ is a *3-dimensional rigid-analytic subvariety* ? Is it a rigid analytic subspace of affine three-space?

There are also certain foundational questions which deserve much better understanding than we have, at present. For example, one still lacks a satisfactory conceptual definition of what it means for a modular form (of general p-adic weight) to be *overconvergent.* This notion is clear for integral weight, but at present, for general weight, we have only an ad hoc procedure based on the availability of families of Eisenstein series of general weight (See Chapter 2 below: you multiply your modular form by such an Eisenstein series to get the product to be of weight 0 and then ask that this be overconvergent as a rigid analytic function on a suitable affinoid in the modular curve). As a consequence of this awkward strategy, treatment of any serious property about overconvergent modular forms depends upon our detailed understanding of the family of Eisenstein series. It would be good to

a. have a more direct definition of overconvergent modular forms and of families of overconvergent modular forms, and at the same time

b. have a closer understanding of the properties of families of Eisenstein series (including knowledge of their zero-free regions).

An improvement in our state of knowledge of **b** would help in our understanding of the detailed geometry of the eigencurve.

It would be important, as well, to have some detailed computations of specific affinoid subdomains of C_p. In this regard, see forthcoming work of Matthew Emerton who (augmenting earlier calculations of Coleman and Teitelbaum) gives a complete description of the geometry of that part of the eigencurve for $p = 2$ and tame level 1 having minimal slope for their weight (cf. [Em]). See also [CTS] where some results on the low slope part of the 3-adic eigencurve are proven. We might mention here that in these examples ($p = 2, 3$), the slope tends to zero as one approaches the "boundary" of

weight space. Are there components of the p-adic eigencurve (for some p) where this phenomenon does not occur?

To be sure, a satisfactory general theory of the eigencurve must deal with *all* tame levels N (and as a prelude for this, one necessary task, which we carry out, is to set up the deformation theory of pseudo-representations with fixed tame level). In the present paper, although we make the construction of the eigencurve only for $N = 1$, in some sections (where it is easy to do so) we work with more general level N in preparation for the general theory. Our conventions concerning level will be signaled at the start of each section.

Is there an a priori deformation-theoretical approach to the eigencurve, and to the Galois representations that the eigencurve parameterizes? (We explain more precisely what we mean by this at the end of Chapter 1.) Our lack of such an approach accounts for some of the difficulty we have in analyzing local properties of the eigencurve. Is there a natural formal scheme over $\mathbf{Z}_p$ whose associated rigid analytic space is the eigencurve? One is, in any event, guaranteed (by the preprint [LvP]) that any connected one-dimensional, separated, rigid analytic space over $\mathbf{Q}_p$ is the generic fiber of some formal scheme which is flat over $\mathbf{Z}_p$. It might be interesting to study irreducible components of the closed fiber of a formal scheme whose generic fiber is (a piece of) the eigencurve. In this connection, J. Teitelbaum has a computer program that produces such irreducible components in a range of slopes (for $p = 2, 3$).

In a later publication we hope to present more foundational material regarding the connection between Katz modular function and convergent eigenforms (including the proof of the compatibility of the action of the diamond operators, i.e., the proof of Prop. 3.4.2 below which we omitted from this article). We are deeply indebted to Kevin Buzzard for his extremely helpful comments throughout the preparation of this article. We also wish to thank Brian Conrad for his helpful suggestions on Chapter 1 and Matthew Emerton for his close reading of an early draft.

Table of Contents.

Chapter 1. Rigid analytic varieties.

1.1 Rigid analytic spaces attached to complete local noetherian rings.

Let p be a prime number, and $\mathbf{C}_p$ the completion of a fixed algebraic closure of $\mathbf{Q}_p$. Let v be the continuous valuation on $\mathbf{C}_p$ such that $v(p) = 1$. Define the absolute value $|\ |$ by $|a| = p^{-v(a)}$, for $a \in \mathbf{C}_p$. By the ring of integers $\mathcal{O}_K$ in a complete subfield K of $\mathbf{C}_p$ we mean the set of elements of absolute value at most 1. For a rigid analytic variety Y (see §9.3 of [BGR]), $A(Y)$ will denote the ring of rigid functions on Y, for $a \in A(Y)$, $|a|$ will denote the spectral semi-norm of a (which may be infinite valued) and $A^0(Y)$ will denote the sub-$\mathcal{O}_K$-algebra of rigid functions with spectral semi-norm at most 1. For an affinoid algebra B, we let $\mathrm{Max}(B)$ denote the corresponding rigid variety over K.

Now let $R = \mathcal{O}_K$ be the ring of integers in a fixed finite extension K of $\mathbf{Q}_p$ in $\mathbf{C}_p$. Let k denote the residue field of R. Let A be a complete local noetherian R-algebra with maximal ideal m_A and residue field $A/m_A = k$. Consider the functor $A \longrightarrow X_A$ which attaches to each such complete noetherian local ring, its associated rigid analytic space over K. We refer to section 7 of [de J] for the construction, and for its basic properties (cf. loc. cit. Definition 7.1.3, 7.1.4a, and 7.1.5). Briefly, the construction may be given as follows: For each real number $r > 0$, let A_r be the p-adic completion of the quotient by its p-power torsion, of the ring

$$A[\{y_{r,(a)}\}]$$

where (a) ranges over unordered tuples of elements in m_A subject to the relations

$$p^{[k(a)r]} y_{r,(a)} = \prod (a),$$