# CONGRUENCES BETWEEN MODULAR FORMS

FRANK CALEGARI

Contents



## 1. Basics

1.1. **Introduction.** The theory of modular forms — and the numerous congruence properties that their coefficients enjoy — can be approached on many levels. Take,

for example, the following congruence of Ramanujan [Ram16]:

$$\Delta := q\prod_{n=1}^{\infty}(1-q^n)^{24} \equiv \sum_{n=1}^{\infty}\sigma_{11}(n)q^n \mod 691.$$

To prove this congruence requires knowing only three facts: that both $\Delta$ and $E_{12}$ are classical modular forms of weight 12, that the ring of classical modular forms is given by $\mathbf{Z}[E_4, E_6] \otimes \mathbf{C}$, and that the numerator of $B_{12}$ is divisible by 691. At the same time, this congruence also points towards a deeper strucure; it represents the first incarnation of the main conjecture of Iwasawa theory — a theorem relating the special values of Dirichlet $L$-functions to corresponding eigenspaces of class groups of abelian extensions of $\mathbf{Q}$. The theory of congruences of modular forms can be (roughly) distinguished into two types:

(1) congruences between Hecke *eigenforms*,
(2) congruences between classical holomorphic or meromorphic modular forms.

The first subject is very rich indeed and encompasses (broadly construed) the entire theory of two dimensional odd Galois representations of $G_{\mathbf{Q}}$. We shall not concern ourselves with such congruences in these notes (except to the extent that they are required to understand congruences of the second kind). Instead, we shall grapple with the second class of congruences, which has as its genesis various conjectures of Ramanujan concerning the partition function proved by Watson [Wat38] (see Theorem 1.1.2 below). Throughout this text, we shall consider the following two examples, which, although enjoying some special properties which distinguish them slightly from the general case, exhibit the typical behavior with respect to the type of congruences treated in these notes.

Let

$$j = \frac{\left(1+240\sum\sigma_3(n)q^n\right)^3}{q\prod_{n=1}^{\infty}(1-q^n)^{24}} = \frac{1}{q} + 744 + 196884q + \ldots =: \sum c(n)q^n.$$

This is the function known as Klein's modular invariant (or, in a slightly different context, as simply the $j$-invariant). It is a meromorphic modular function of weight zero, and is the unique such function which is holomorphic away from a simple pole at the cusp such that $j(\rho) = 0$ and $j(i) = 1728$. The $q$-expansion $j$ has coefficients in $\mathbf{Z}$ which grow sub-exponentially but faster than polynomially. We shall be interested in the congruence properties of the coefficients $c(n)$.

1.1.1. **Exercise.** The Wikipedia entry on the $j$-invariant is embarrassing — make it better.

Our second example (which we consider more briefly) will concern the inverse of Dedekind's eta function, which is (essentially) the generating function for the partitions, namely:

$$\eta^{-1} = \frac{1}{q^{1/24}\prod_{n=1}^{\infty}(1-q^n)} = \sum p\left(\frac{n+1}{24}\right)q^n.$$

(Following the standard convention, $p(m) = 0$ if $m$ is not an integer.)

The following theorems are due to Lehner [Leh49] and Watson [Wat38] respectively.

1.1.2. **Theorem.** *The following congruences are satisfied:*

(1) *If* $n \equiv 0 \mod 2^m$ *and* $n > 0$, *then* $c(n) \equiv 0 \mod 2^{3m+8}$.
(2) *If* $24n \equiv 1 \mod 5^m$, *then* $p(n) \equiv 0 \mod 5^m$.

The proof of these congruences relied on a judicious use of *modular equations*, that is, the explicit functional relationships between modular functions of certain small levels. We shall thereby dub this technique the *modular equation method*; it was also applied by Atkin and O'Brien to prove similar congruences for higher moduli [AO67]. We shall argue that the systematic use of overconvergent modular forms is *a direct descendant of the modular equation method*.

1.1.3. *The scope of this document.* These notes are *not* intended to be an introduction to the theory of modular forms, although we shall summarize some of the salient details. Rather, it is directed towards three specific audiences, namely:

(1) Graduate students in number theory with a basic understanding of classical modular forms and their $q$-expansions.
(2) Those who are interested in congruences concerning specific modular forms, for example involving partitions, but whom are not fully conversant with the modern geometric and rigid analytic viewpoint of Dwork, Katz, and Coleman.
(3) Those who understand the theory of overconvergent modular forms, and are curious about the applications to concrete congruences.

Since these audiences by definition have somewhat different backgrounds, I will have to apologize in advance for saying things that you, dear reader, will find obvious. I will also apologize for eliding technical details whose absence may push the more careful reader into an apoplectic fit. However, the theory of elliptic curves and modular forms encompasses quite a lot of mathematics, and so I will necessarily be cursory on several important points (most importantly, the technical details concerning the construction of modular curves [DR73, KM85], as well as any rigorous details at all concerning rigid analytic spaces). In particular, I will concentrate on the issues that are most relevant to my purpose, and leave the secondary matter to the literature, which is extensive and (quite frequently) very well written, e.g. [Sil86, Sil94, Kat73, DS05, Buz03]. Indeed, as with any lecture notes, the key choice is to decide which points to elide, which points to skip, and which points to emphasize. Since much of what I say in the first half of these notes overlaps with what is in [Kat73], I leave out several arguments that Katz gives in detail, and instead concentrate on giving examples and emphasizing the points that some might find confusing if approaching [Kat73] with limited background. Let me include at this point the following table, whose content[1] is self-explanatory.

| The modern method | The classical antecedent |
|---|---|
| The compactness of the $U$ operator | The modular equations method |
| Serre weights and Ash–Stevens | The weight filtration in low weights |
| $p$-adic Langlands for $\mathrm{GL}_2(\mathbf{Q}_p)$ | The weight filtration in higher weights |
| Holomorphic sections over the ordinary locus | Serre's $p$-adic modular forms |

[1]For reasons of time, I will not discuss in any detail the second and third rows of this table. For the connection between the $\theta$-operator and Serre weights, one should consult [AS86]. The only time these ideas arise in any form within these notes is secretly — via an appeal to a result of Buzzard–Gee [BG09] concerning Galois representations associated with small slope forms. However, I will suppress all of the details of that paper, together with their concomitant difficulties relating to, *inter alia*, $p$-adic local Langlands.

I make no claim to the originality of most of the material presented here. Many of the ideas here can be found in Katz [Kat73] and Coleman [Col96, Col97]. The general philosophy regarding asymptotic expansions is in Gouvea–Mazur [GM95]. The modular equation method of [Wat38] has its roots firmly in 19th century mathematics. I learned many of these ideas through conversations with Matthew Emerton that started in 1993, and with Kevin Buzzard which started in 2001 — in relation to his Arizona Winter School project. One reason I decided to write on this particular topic was that, being conversant with some of the explicit aspects of theory, I might be more in a position than most to bridge the divide between the classical and modern perspectives on congruences. Thanks go to Simon Marshall for some conversations about adjoint $L$-functions and arithmetic quantum unique ergodicity that influenced some of the wild conjectures of this paper, thanks also to David Loeffler for some conversations and for making available some of his previous computations. Thanks to Rebecca Bellovin, George Boxer, Ana Caraiani, Martin Derickx, Toby Gee, James Newton, and David Savitt for some corrections, and thanks to Matt Baker for some helpful remarks concerning $p$-adic equidistribution.

1.1.4. *A note on the exercises.* Some of the exercises are easy, some are tricky, most I know how to do, but some I do not. I put a $\star$ on the particularly tricky exercises.

1.2. **What is a modular form?** There are many (more or less general) definitions of a modular form. A good source (which we follow here) is Katz [Kat73]. The classical definition, which for most purposes is not particularly useful, is that a modular form $f$ of weight $k$ for $\mathrm{SL}_2(\mathbf{Z})$ is a holomorphic function on the upper half plane $\mathbf{H}$ satisfying the functional equation

$$f\left(\frac{a\tau+b}{c\tau+d}\right) = (c\tau+d)^k f(\tau)$$

for all $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z})$, and such that $f(\tau)$ is bounded as $\tau \to i\infty$. A slightly more useful definition is the following:

1.2.1. **Definition** (Version 1)**.** *A modular form $f$ of weight $k$ over $\mathbf{C}$ is a function on lattices $\Lambda = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2 \subset \mathbf{C}$ such that:*

(1) *$f(\mathbf{Z}\tau + \mathbf{Z})$ is holomorphic as a function of $\tau$,*
(2) *$f(\mu\Lambda) = \mu^{-k} f(\Lambda)$ for all $\mu \in \mathbf{C}^\times$,*
(3) *$f(\mathbf{Z}\tau + \mathbf{Z})$ is bounded as $\tau \to i\infty$.*

We say that two lattices $\Lambda$ and $\Lambda'$ are *homothetic* exactly when there exists a $\mu \in \mathbf{C}^\times$ such that $\Lambda = \mu\Lambda'$.

The following theorem is proved in almost any book on modular forms or elliptic curves.

1.2.2. **Theorem** (Weierstrass)**.** *Given a lattice $\Lambda = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$, the quotient $E := \mathbf{C}/\Lambda$ has the structure of a smooth projective curve of genus one given by the (affine) equation*

$$E : y^2 = 4x^3 - 60G_4 x - 140G_6,$$

*where*

$$G_4 = \sum_{\Lambda\setminus 0} \frac{1}{\lambda^4}, \qquad G_6 = \sum_{\Lambda\setminus 0} \frac{1}{\lambda^6}.$$

*Explicitly, the map is given by*

$$x = \wp(z,\tau) = \frac{1}{z^2} + \sum_{\Lambda\setminus 0} \frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2}, \qquad y = \frac{dx}{dz} = \wp'(z,\tau) = \sum \frac{-2}{(z-\lambda)^3}.$$

*Moreover, every elliptic curve over* $\mathbf{C}$ *admits such a uniformization.*

In particular, given a lattice $\Lambda$, one obtains an elliptic curve $E$. (An elliptic curve $E$ over $\mathbf{C}$ is, by definition, a smooth genus one curve with a marked point, which in this case is the point "at infinity", e.g. $z = 0$. Showing that elliptic curves over $\mathbf{C}$ admit a Weierstrass equation is an elementary exercise using the Riemann–Roch Theorem.) It is important to note, however, this map is not a bijection. It is easy to see that if one *scales* the lattice $\Lambda$ by a homothety, say replacing $\Lambda$ by $\mu\Lambda$, then $G_4$ is replaced by $\mu^{-4}G_4$ and $G_6$ is replaced by $\mu^{-6}G_6$. The corresponding elliptic curves are isomorphic under a scaling in $x$ and $y$. In particular, this map is a bijection between lattices $\Lambda$ in $\mathbf{C}$ *up to homothety* and elliptic curves $E$ over $\mathbf{C}$. Modular forms, however, are not functions on lattices up to homothety unless $k = 0$. It is natural to ask, therefore, whether Weierstrass' theorem gives a natural bijection between lattices and elliptic curves enriched with some extra structure.

1.2.3. **Lemma.** *The space of holomorphic differentials on an elliptic curve* $E$ *over* $\mathbf{C}$ *is one dimensional, that is,* $H^0(E,\Omega^1) = \mathbf{C}$.

*Proof.* If one defines an elliptic curve to be a smooth projective curve of genus one, then this lemma is a tautology. If one imagines an elliptic curve to be given by a quotient $E = \mathbf{C}/\Lambda$, then one can argue as follows. Any holomorphic differential pulls back to a differential $\omega = f(z)dz$ on $\mathbf{C}$ which is invariant under translation. Since the differential $dz$ has no poles and no zeroes, it follows that $f(z)$ must be holomorphic on $\mathbf{C}$ and doubly periodic, and thus (by Liouville's theorem) constant. Hence the only such differential (up to scalar) is $dz$. □

This definition allows us to understand what *extra structure* a lattice contains beyond the isomorphism class of the corresponding elliptic curve.

1.2.4. **Lemma.** *There is a bijection between lattices* $\Lambda \subset \mathbf{C}$ *and elliptic curves* $E$ *together with a non-zero differential* $\omega \in H^0(E,\mathbf{C})$. *The bijection is given by taking a lattice* $\Lambda$ *to the corresponding Weierstrass equation, and then taking*

$$\omega := dz = \frac{dx}{dx/dz} = \frac{dx}{y}.$$

This bijection is canonical, but it is not *canonically* canonical, since one could form such a bijection using any fixed multiple of $dX/Y$. This choice, however, also makes sense *integrally* (at least up to factors of 2). Under this bijection, we can compute explicitly what happens if we replace $\Lambda$ by $\mu\Lambda$. We do this in gory detail. Explicitly:

$$\Lambda \mapsto\colon y^2 = 4x^3 - 60G_4x - 140G_6, \qquad \omega_\Lambda = \frac{dx}{y},$$

$$\mu\Lambda \mapsto\colon v^2 = 4u^3 - 60G_4\mu^{-4}u - 140G_6\mu^{-6}, \qquad \omega_{\mu\Lambda} = \frac{du}{v}.$$

Now we may write the latter curve as $(v\mu^3)^2 = 4(u\mu^2)^3 - 60G_4(u\mu^2) - 140G_6$, and hence

$$\omega_{\mu\Lambda} = \frac{du}{v} = \frac{dx\mu^{-2}}{y\mu^{-3}} = \mu \cdot \omega_\Lambda.$$

This leads to a new definition of a modular form.

1.2.5. **Definition** (Version 2)**.** *A modular form $f$ of weight $k$ over $\mathbf{C}$ is a function on pairs $(E, \omega)$ consisting of an elliptic curve $E$ and a non-zero element $\omega \in H^0(E, \Omega^1_E)$ such that*

$$f(E, \mu\omega) = \mu^{-k} f(E, \omega),$$

*and such that $f(\mathbf{C}/(\mathbf{Z}\tau + \mathbf{Z}), dz)$ is bounded as $\tau \to i\infty$.*

This coincides with the previous definition if we let $f(\Lambda) = f(\mathbf{C}/\Lambda, dz)$, since (as we saw above) $f(\mu\Lambda) = f(\mathbf{C}/\Lambda, \mu dz)$.

1.2.6. *The fundamental domain and $X(1)$.* The action of $\mathrm{PSL}_2(\mathbf{Z})$ on the upper half plane $\mathbf{H}$ has a fundamental domain given by the shaded region in Figure 1.2.6:

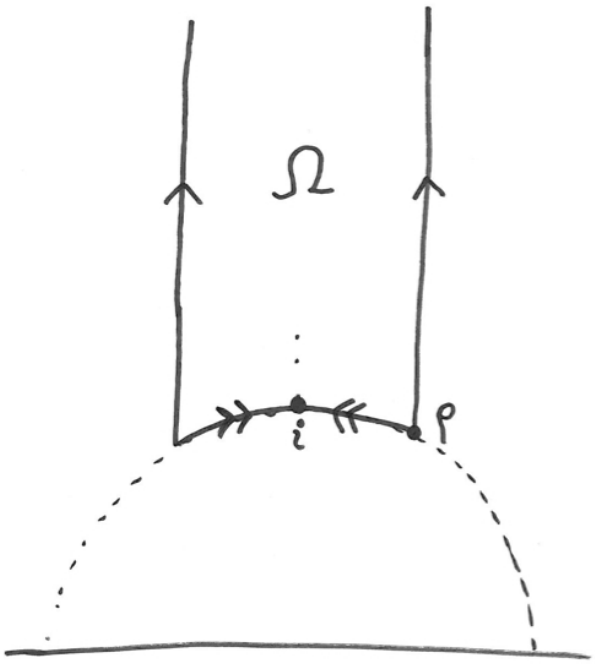


FIGURE 1. This is a picture you have seen before

The quotient $Y(1) = \mathbf{H}/\mathrm{PSL}_2(\mathbf{Z})$ has a natural structure as a complex orbifold with cone points of angles $2\pi/2$ at $z = i$, $2\pi/3$ and $z = \rho$, and $2\pi/\infty$ at $z = i\infty$. There is a natural compactification $X(1)$ obtained by filling in the cusp at infinity. The function

$$j : X(1) \to \mathbf{P}^1$$

is a bijection over the complex numbers, and thus can (roughly) be interpreted as giving $X(1)$ the structure of the complex variety $\mathbf{P}^1_j$ with $j$ as a uniformizing parameter. On the other hand, the corresponding map $j : \mathbf{H} \to \mathbf{P}^1$ has the property that $j$ has a triple zero at $z = \rho$ and $j - 1728$ has a double zero at $z = i$. In particular, the functions $\sqrt[3]{j}$ and $\sqrt{j - 1728}$ extend to holomorphic functions on $\mathbf{H}$ (they are no longer invariant under $\mathrm{PSL}_2(\mathbf{Z})$, although they are modular functions for various congruence subgroups). The two ways of thinking about $j$ reflect the difference between the underlying *orbifold* structure and the topological structure, which ultimately is related to the fact that $X(1)$ — as a moduli space — is more properly thought of as a stack. However, thinking in terms of stacks is not at all necessary in this case, by virtue of the fact that these issues can (essentially) be completely avoided by working at higher level.

1.2.7. *Modular forms with level structure.* It is natural to consider modular forms for various special *subgroups* of $\mathrm{SL}_2(\mathbf{Z})$. The most natural ones are defined as follows.

1.2.8. **Definition.** *For an integer $N$, the congruence subgroups $\Gamma_0(N)$, $\Gamma_1(N)$, and $\Gamma(N)$ are defined as follows:*

$$\Gamma_0(N) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; \gamma \in \mathrm{SL}_2(\mathbf{Z}), \gamma \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \mod N \right\}$$

$$\Gamma_1(N) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; \gamma \in \mathrm{SL}_2(\mathbf{Z}), \gamma \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \mod N \right\}$$

$$\Gamma(N) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; \gamma \in \mathrm{SL}_2(\mathbf{Z}),\ \gamma \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod N \right\}$$

These groups act naturally on $\mathbf{H}$ via the action of $\mathrm{SL}_2(\mathbf{Z})$. The sets $\mathbf{H}/\Gamma$ are naturally in bijection with the following sets:

(1) $\mathbf{H}/\Gamma_0(N)$ is naturally in bijection the following set up to homothety: Pairs $(\Lambda, \Sigma)$ consisting of a lattice $\Lambda$ together with a cyclic subgroup $\Sigma \subset \mathbf{C}/\Lambda$ of order $N$.
(2) $\mathbf{H}/\Gamma_0(N)$ is naturally in bijection following set up to homothety: Pairs $(\Lambda, \Lambda')$ consisting of a pair of lattices $\Lambda$, $\Lambda'$ with $\Lambda \subset \Lambda'$ and $\Lambda'/\Lambda \simeq \mathbf{Z}/N\mathbf{Z}$.
(3) $\mathbf{H}/\Gamma_1(N)$ is naturally in bijection with the following set up to homothety: Pairs $(\Lambda, P)$ consisting of a lattice $\Lambda$ together with a point $P \in \mathbf{C}/\Lambda$ of exact order $N$.
(4) $\mathbf{H}/\Gamma(N)$ is naturally in bijection the following set up to homothety: Lattices $\Lambda$ together with a commutative diagram:

$$\begin{array}{ccc} \frac{1}{N} \cdot \Lambda/\Lambda & \xrightarrow{\simeq} & (\mathbf{Z}/N\mathbf{Z})^2 \\ \wedge \downarrow & & \downarrow \wedge \\ \mu_N & = & \mathbf{Z}/N\mathbf{Z} \end{array}$$

where $\wedge$ is the Weil pairing on the left hand side and the symplectic pairing $(\mathbf{Z}/N\mathbf{Z})^2 \to \mathbf{Z}/N\mathbf{Z}$ given by $(a, b) \wedge (c, d) = ad - bc$ on the right hand side, and $\mu_N \to \mathbf{Z}/N\mathbf{Z}$ is the map sending a fixed root of unity $\zeta_N$ to 1.

The sets (1) and (2) are in bijection. If $\pi$ denotes the projection $\pi : \Lambda \mapsto \Lambda/N\Lambda$, then

$$(\Lambda, \Sigma) \mapsto (\Lambda, \pi^{-1}(\Sigma)), \qquad (\Lambda, \Lambda') \mapsto (\Lambda, \pi(\Lambda')).$$

One might wonder why the set $\mathbf{H}/\Gamma(N)$ is not simply in bijection with lattices $\Lambda$ together with an isomorphism $\Lambda/N\Lambda \mapsto (\mathbf{Z}/N\mathbf{Z})^2$. The reason is the space of such pairs is not connected in the natural topology: the Weil pairing of any chosen basis $(P, Q)$ of $\Lambda/N\Lambda$ yields (locally) a continuous map from this space to $\mu_N$, which is thus constant. Moreover, if one passes from one lattice to an equivalent one (both given by a point $\tau$ in the upper half plane), the corresponding change of basis matrix lies in $\mathrm{SL}_2(\mathbf{Z})$, and thus the effect on the chosen basis $(P, Q)$ is via an element of the image:

$$\mathrm{SL}_2(\mathbf{Z}) \mapsto \mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z}).$$

Yet this map is not surjective — the image consists exactly of the matrices of determinant $+1$. In particular, the locally constant function extends to a global function from this space to $\mu_N$. In particular, this space is most naturally isomorphic to

$$|(\mathbf{Z}/N\mathbf{Z})^\times|$$

copies of $\mathbf{H}/\Gamma(N)^2$.

1.2.9. *Modular curves as complex manifolds.* If $\Gamma$ is a finite index subgroup of $\mathrm{SL}_2(\mathbf{Z})$, we let $Y(\Gamma)$ denote the quotient $\mathbf{H}/\Gamma$. If one lets $\mathbf{H}^* = \mathbf{H} \cup \mathbf{P}^1(\mathbf{Q})$ (where $\infty$ corresponds to $i\infty$) then $\mathbf{H}^*$ admits a natural action of $\mathrm{SL}_2(\mathbf{Z})$, and $\mathbf{H}^* \setminus \mathbf{H}$ is a single orbit. The quotients $X(\Gamma) := \mathbf{H}^*/\Gamma$ provide natural compactifications of $Y(\Gamma)$. If $\Gamma$ is torsion free, then $X(\Gamma)$ and $Y(\Gamma)$ are smooth complex manifolds, and indeed $Y(\Gamma)$ is a $K(\pi, 1)$-space with $\pi_1(Y(\Gamma)) = \Gamma$. If $\Gamma = \Gamma_0(N)$, $\Gamma_1(N)$, or $\Gamma(N)$, we write $X_0(N)$, $X_1(N)$, and $X(N)$ respectively for the corresponding spaces. If $\Gamma' \subset \Gamma$ has finite index, the natural map:

$$X(\Gamma') \to X(\Gamma)$$

is smooth away from the cusps and the preimages of $i$ and $\rho$ in $X(1)$. Computing the genus of modular curves is a simple exercise from Galois theory and from the Riemann–Hurwitz formula.

1.2.10. **Example.** *Let $p$ be prime. The genus of $X(p)$ is given by*

$$g(X(p)) = \begin{cases} 0, & p = 2 \\ \dfrac{(p+2)(p-3)(p-5)}{24}, & p > 2. \end{cases}$$

Consider the case when $p \geq 3$ for convenience. The map $X(p) \to X(1)$ is a Galois covering which — in an orbifold sense —is smooth away from the cusps. However, thinking of $X(1)$ as $\mathbf{P}^1$, there will be ramification above $i$ and $\rho$ of degree 2 and 3 respectively (since $X(p)$ is a manifold for $p \geq 3$). We may thus use the Riemann–Hurwitz formula. Since $X(1)$ has only one cusp, the group $G = \mathrm{PSL}_2(\mathbf{F}_p)$ acts

[2]As seen below, we denote the corresponding algebraic curves (with complex points $\mathbf{H}/\Gamma(N)$ or $\mathbf{H}^*/\Gamma(N)$) by $Y(N)$ and $X(N)$ respectively. However, there are alternate definitions of these curves which are *not* geometrically connected. Since one usually uses the same notation, one distinguishes them by talking about the *big* $X(N)$ and the *small* $X(N)$.

transitively on the cusps of $X(p)$. The stabilizers of the cusps are all isomorphic, but the stabilizer of $\infty$ is clearly equal to

$$\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix},$$

which is a group of order $p$. Hence, by the Orbit–Stabilizer theorem, the number of cusps is equal to

$$c = \frac{1}{p} \cdot |G|.$$

Similarly, there are $|G|/2$ and $|G|/3$ points above $i$ and $\rho$ respectively. Hence, by Riemann–Hurwitz, we have

$$\begin{aligned} \chi(X(p)) &= |G|\chi(X(1)) - \frac{(2-1)|G|}{2} - \frac{(3-1)|G|}{3} - \frac{(p-1)|G|}{p}, \\ &= |G| \left( 2 - \frac{1}{2} - \frac{2}{3} - 1 + \frac{1}{p} \right) = -\frac{(p^2-1)(p-6)}{12}, \end{aligned}$$

and hence

$$g(X(p)) = \frac{(p+2)(p-3)(p-5)}{24}.$$

A similar calculation works for $X_0(p)$ and $X_1(p)$ — although the covers are no longer Galois in these cases, and the ramification at $i$ and $\rho$ depends on the reduction of $p$ modulo 12. Note that the action of $\Gamma_0(p)$ on $\mathbf{P}^1(\mathbf{Q})$ has two orbits, corresponding to $\infty$ and 0. The cusp $\infty$ is unramified, and the cusp 0 is ramified of degree $p$.

1.2.11. *Modular curves as algebraic curves.* The curves $X(\Gamma) = \mathbf{H}^*/\Gamma$ are compact Riemann surfaces, and so, by a theorem of Riemann, are algebraic curves.

1.2.12. **Exercise.** *Why are compact complex manifolds of dimension one algebraic? Understand why the key point is the existence on $X$ of a meromorphic differential $\omega$. Also, understand why the result fails in higher dimensions.*

1.2.13. *Modular curves as moduli spaces.* Another way to define the modular curves $Y(\Gamma)$ for suitable $\Gamma$ is to define them as *moduli spaces* for appropriately defined functors. For example, $Y_1(N)$ is the fine moduli space for pairs $(E, P)$ of elliptic curves $E$ together with an inclusion of group schemes $\mathbf{Z}/N\mathbf{Z} \to E$. For $Y(N)$, is is natural to consider elliptic cuves $E$ together with an isomorphism $\mathbf{Z}/N\mathbf{Z} \oplus \mu_N \to E[N]$ (although this leads to the "big" $Y(N)$ rather than the small one). There are technical issues relating to this construction for $\Gamma = \Gamma_0(N)$ due to the presence of automorphisms, but for $Y_1(N)$ and $Y(N)$ (at least for $N > 4$ and $N \geq 3$ respectively) it provides a construction of the appropriate spaces as algebraic varieties. By definition, the curves $Y_1(N)$ and $Y(N)$ then come along with a universal modular curve $\mathcal{E}/Y(\Gamma)$. There are several ways one might like to improve this construction:

(1) Make the construction more arithmetic, so it defines a smooth curve over $\mathbf{Q}$ (or $\mathbf{Q}(\zeta_N)$), or even the integral rings $\mathbf{Z}[1/N]$ and $\mathbf{Z}[1/N, \zeta_N]$. Even better, construct a nice model over $\mathbf{Z}$ with good reduction over $\mathbf{Z}[1/N]$.
(2) Extend the construction in a natural way to the cusps.
(3) Do this all in a moduli theoretic way.

That this can be done is not an entirely trivial proposition [DR73, KM85]. On the other hand, it is not hugely complicated either, and one is "lucky" (compared to the moduli space of higher dimensional abelian varieties, say) that the cusps are not all that complicated in the end. For our purposes we can take these on faith.

We may now extend our definition of a modular form to general rings $R$.

1.2.14. **Definition** (Version 2$a$)**.** *A meromorphic modular form $f$ of weight $k$ over $R$ is a function on pairs $(E/A, \omega)$ where $\omega$ is a nowhere vanishing section of $\Omega^1_{E/A}$ and $A$ is an $R$-algebra such that:*

(1) *$f(E/A, \omega)$ depends only on the $A$-isomorphism class of $(E/A, \omega)$.*
(2) *$f(E, \mu\omega) = \mu^{-k} f(E, \omega)$ for any $\mu \in A^\times$.*
(3) *If $\phi : A \to B$ is any map of rings, then $f(E/B, \omega_B) = \phi(f(E/A, \omega))$.*

One deficit with this definition is that it doesn't address the issue at the cusps, in order to address this we will need to say something about Tate curves.

1.2.15. **Exercise.** *Let $R$ be a ring in which 6 is invertible. Prove that, given a pair $(E/R, \omega)$, there exists a Weierstrass equation:*

$$y^2 = x^3 + a_4 x + a_6$$

*for $E$ such that*

$$\omega = dx/y.$$

*Prove that the rules $f(E/R, \omega) = a_4$ and $g(E/R, \omega) = a_6$ define modular forms of weights 4 and 6 respectively.*

1.2.16. *Modular forms and modular functions on $X(\Gamma)$.* Naturally enough, the various definition of modular forms of level one each extend to corresponding definitions in higher weight. For any fixed level $\Gamma$, we consider the points on $\mathbf{H}/\Gamma$ as pairs $(\Lambda, \alpha)$ up to homothety, where $\alpha$ denotes the extra structure. One defines modular forms of higher weight simply to be rules on triples $(E/R, \omega, \alpha)$ which are compatible in the natural way.

1.2.17. *Modular forms as sections of a line bundle.* How does $H^0(E, \Omega_E)$ vary as one winds around the curve $Y(\Gamma) = \mathbf{H}/\Gamma$? If we start with a curve $E_\tau = \mathbf{C}/\{\mathbf{Z} + \tau\mathbf{Z}\}$, together with its canonical differential $dz$, then we can imagine moving (in the upper half plane) from $\tau$ to

$$\gamma\tau = \tau' = \frac{a\tau + b}{c\tau + d}, \qquad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$$

which brings us back to the same elliptic curve $E_{\tau'} \simeq E_\tau$. The invariant differential $dz \in H^0(E, \Omega_E)$ varies continuously as we vary $E$, yet when we return to $E$ we observe that $\omega_E$ and $\omega_{\gamma E}$ have changed, in particular,

$$dz \in H^0(\mathbf{C}/\{(a\tau + b)\mathbf{Z} + (c\tau + d)\mathbf{Z})\}, \Omega^1) \longleftrightarrow (c\tau + d)dz \in H^0(\mathbf{C}/\{\mathbf{Z} + \tau'\mathbf{Z}\}, \Omega^1).$$

In particular, the behavior of $(dz)^{\otimes k}$ as one winds around $Y(\Gamma)$ via the element $\gamma$ exactly corrects the corresponding behavior of a modular form of weight $k$. This leads to the identification of modular forms (ignoring issues at the cusps) as sections of some line bundle $\mathcal{L}$ whose fibers at a point $E$ are naturally isomorphic to $H^0(E, \Omega^1_E)^{\otimes k}$. By "naturally", we mean that the monodromy of this bundle is as computed above. How may one construct such a line bundle? We want to "interpolate" the (trivial) sheaf $\Omega^1$ as $E$ varies over $Y(\Gamma)$. To do this, one can consider

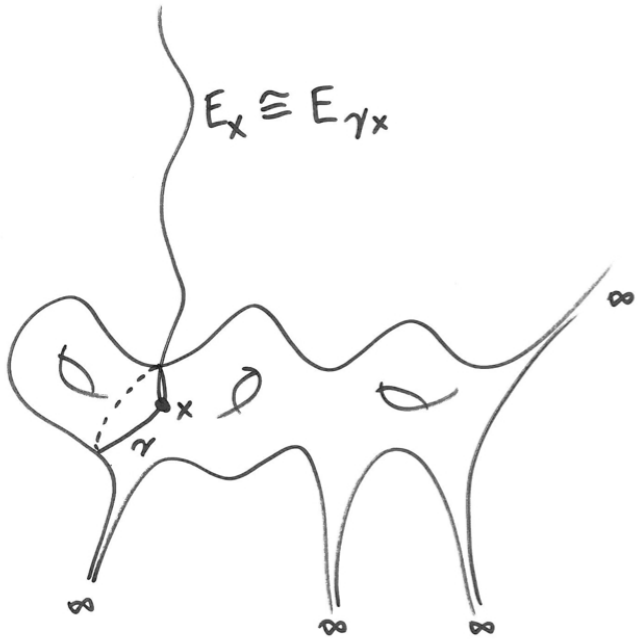


FIGURE 2. The monodromy of $\omega_X$ for $\gamma \in \Gamma$

the sheaf of *relative differentials* $\Omega^1_{\mathcal{E}/Y(\Gamma)}$ on $Y(\Gamma)$. If $\pi : \mathcal{E} \to Y(\Gamma)$ denotes the natural projection, then we let:

$$\omega_Y := \pi_* \Omega^1_{\mathcal{E}/Y}.$$

Intuition tells us that the fibre of $\omega_Y$ at a point $E \in Y$ corresponding to an elliptic curve $E$ should be exactly what we are looking for, i.e.,

$$\Omega^1_E = H^0(E, \Omega^1_E)$$

(a sheaf at a closed point is simply the module of global sections). This intuition is correct — it requires only that the map $\pi$ is proper.

1.2.18. **Definition** (Version 3*a*)**.** *A meromorphic modular form $f$ of weight $k$ over $R$ and level $\Gamma$ is a section of $H^0(Y(\Gamma)_R, \omega^{\otimes k})$.*

Note that we also would like to understand what happens at the cusps. Fortunately, the construction of [DR73] provides us with a generalized elliptic curve $\mathcal{E}/X(\Gamma)$, and a corresponding line bundle $\omega_X$ on $X(\Gamma)$, and we may set:

1.2.19. **Definition** (Version 3*b*)**.** *A modular form $f$ of weight $k$ over $R$ and level $\Gamma$ is a section of $H^0(X(\Gamma)_R, \omega^{\otimes k})$.*

It's usually sensible to assume that $R$ is a $\mathbf{Z}[1/N]$-algebra where $N$ is the level of $\Gamma$. Denote the $R$-module $H^0(X(\Gamma)_R, \omega^{\otimes k})$ of modular forms by $M_k(\Gamma, R)$.

1.2.20. **Warning.** *In order to define $\omega_Y$ or $\omega_X$, one needs the existence of a universal generalized elliptic curve $\mathcal{E}$, which requires the moduli problem to be fine, which*

*requires working with $X_1(N)$ rather than $X_0(N)$. This is not an artificial problem — there is no appropriate sheaf $\omega$ on $X_0(N)$, and modular forms of odd weight on $X_0(N)$ are automatically zero for parity reasons.*

1.2.21. **Exercise.** *Show that — when it makes sense to compare them — all the definitions of meromorphic modular forms coincide.*

1.2.22. *Kodaira–Spencer: Another description when $k = 2$.* Another description of modular forms of weight 2 over $\mathbf{C}$ arises from the fact that, for such forms, $f(\tau)d\tau$ is invariant under $\Gamma$. This might lead one to suspect that

$$\Omega^1_X \simeq \omega_X^{\otimes 2},$$

but this is only correct along $Y(\Gamma)$. The problem is that the differential $d\tau$ is not smooth at the cusp. The natural (analytic) parameter at $\infty$ is $q = e^{2\pi i\tau}$, and

$$d\tau = \frac{1}{2\pi i}\frac{dq}{q}.$$

In particular, a section of $H^0(X, \Omega^1)$ will be (locally) a multiple of $dq$, and so the corresponding function $f(\tau)dq = 2\pi i q f(\tau)d\tau$ will automatically vanish at the cusp. In particular, the correct isomorphism is

$$\Omega^1_X(\infty) \simeq \omega^2_X,$$

where $D(\infty)$ indicates that differentials are allowed to have poles of orders at most one at the cusps. These isomorphisms go by the name of the Kodaira–Spencer Isomorphism — over $Y$ it can be deduced more directly using deformation theory (see §3B of [HM98] for a geometric discussion).

1.2.23. *Change of coefficients.* Mostly the coefficients $R$ just come along for the ride. In particular, $M_k(\Gamma, R)$ denotes the forms of weight $k$ and level $\Gamma$ over $R$, then one might hope that

$$M_k(\Gamma, S) = M_k(\Gamma, R) \otimes_R S$$

for an $S$-algebra $R$. This is certainly true if $S$ is a flat $R$-algebra, but it is not *always* true. The exceptions, however, are mainly confined to vary particular circumstances:

1.2.24. **Proposition.** *Let $S$ be an $R$-algebra, and suppose that $N$ is invertible in $R$. Then there is an isomorphism:*

$$M_k(\Gamma(N), S) \simeq M_k(\Gamma(N), R) \otimes_R S$$

*provided that $N \geq 3$ and $k \geq 2$.*

The only interesting case is really when $R = \mathbf{Z}_p$ and $S = \mathbf{F}_p$ for a prime $p$ not dividing $N$. There is a map:

$$0 \to \omega \to \omega \to \omega/p \to 0$$

of line bundles on $X(N)/\mathbf{Z}_p$. Since $j_*X(N)/\mathbf{F}_p \to X(N)/\mathbf{Z}_p$ is a closed immersion, there is an isomorphism

$$H^i(X(N)/\mathbf{F}_p, \omega^{\otimes k}) \simeq H^i(X(N)/\mathbf{Z}_p, j_*\omega^{\otimes k}) = H^i(X(N)/\mathbf{Z}_p, \omega^{\otimes k}/p).$$

It suffices to show that $H^1(X(N)/\mathbf{Z}_p, \omega^{\otimes k}) = 0$. Since this is finitely generated, it suffices to show that $H^1(X(N)/\mathbf{Z}_p, \omega^{\otimes k})/p$ is zero, and hence it suffices to show that $H^1(X(N)/\mathbf{F}_p, \omega^{\otimes k})$ is zero. There is a very natural way to show that the

first cohomology of a locally free line bundle on a smooth curve vanishes, which is to show that the degree of the line bundle is at least $2g-2$ and then to use Riemann–Roch (or Serre duality). In our case, the appropriate estimate for the degree follows from the Kodaira–Spencer isomorphism $\omega^2 \simeq \Omega^1_X(\infty)$ of § 1.2.22 *provided* that $k \geq 2$.

You may wonder where the previous argument used the fact that $N \geq 3$. It comes up in order to even *talk* about the sheaf $\omega$ which requires the universal elliptic curve $\mathcal{E}/X$. In fact, we shall see later that the result is not even true for $N = 1$ — since the Hasse invariant gives a weight 2 form of level 1 modulo 3 which does not lift to characteristic zero.

One may also wonder whether the result is true for $k = 1$. Katz proves that it is true for $N \leq 11$ and leaves open the possibility that it might be true more generally. Yet it is not. In particular, there are modular forms modulo $p$ of weight one which may not lift to characteristic zero.

1.2.25. **Exercise** (Schaeffer)**.** *Show that the map* $M_1(\Gamma_1(7 \cdot 347), \mathbf{Z}_p) \to M_1(\Gamma_1(7 \cdot 347), \mathbf{F}_p)$ *is not surjective when* $p = 935666449040629144864934236346813$*. Or at least, think about* how *one might prove this.*

1.2.26. *Tate Curves.* (See [Sil94]). Let $q = e^{2\pi i \tau}$. The exponential map induces an isomorphism:

$$\mathbf{C}/\mathbf{Z} \oplus \tau\mathbf{Z} \to \mathbf{C}^\times/q^{\mathbf{Z}} = \mathbb{G}_m(\mathbf{C})/q^{\mathbf{Z}}.$$

Writing the Weierstrass parametrization in terms of the parameter $q$ instead of $\tau$, (and changing the scaling by an appropriate factor of $2\pi i$), we find that a model for $\mathbf{C}^\times/q^{\mathbf{Z}}$ is given by ([Sil94])

$$y^2 + xy = x^3 + a_4(q)x + a_6(q),$$

where

$$a_4 = -\sum \frac{n^3 q^n}{1-q^n}, \qquad a_6 = -\sum \frac{(5n^3+7n^5)q^n}{12(1-q^n)}$$

are both in $\mathbf{Z}[[q]]$. The discriminant of this elliptic curve is (of course)

$$\Delta = q\prod_{n=1}^{\infty}(1-q^n)^{24}.$$

In particular, the equation above formally defines an elliptic curve over the Laurent series ring $\mathbf{Z}[[q]][\Delta^{-1}] = \mathbf{Z}((q))$. We call this the *Tate Curve* and denote it by $T(q)$. It provides a description of the universal elliptic curve $\mathcal{E}/X$ over a punctured disc (over $\mathbf{Z}$) at the cusp $\infty$. We may associate to $T(q)$ an canonical differential

$$\omega_{\text{can}} := \frac{dt}{t} \in H^0(T(q), \Omega^1),$$

where $T(q) = \mathbb{G}_m/q^{\mathbf{Z}}$ and $\mathbb{G}_m = \text{Spec}(\mathbf{Z}[t, t^{-1}])$. In particular, given a meromorphic modular form $f$ of weight $k$ (thought of as a rule via definition $2a$), we *define* the $q$-expansion of $f$ to be:

$$f\left(T(q), \omega_{\text{can}}\right) \in \mathbf{Z}((q))$$

1.2.27. **Exercise.** *In what context does the definition* $\mathbb{G}_m/q^{\mathbf{Z}}$ *make sense?*

1.2.28. **Definition** (Version 2*b*)**.** *A modular form $f$ of weight $k$ over $R$ is a function on pairs $(E/A, \omega)$ where $\omega$ is a nowhere vanishing section of $\Omega^1_{E/A}$ and $A$ is an $R$-algebra such that:*

(1) *$f(E/A, \omega)$ depends only on the $A$-isomorphism class of $(E/A, \omega)$.*
(2) *$f(E, \mu\omega) = \mu^{-k} f(E, \omega)$ for any $\mu \in A^\times$.*
(3) *If $\phi : A \to B$ is any map of rings, then $f(E/B, \omega_B) = \phi(f(E/A, \omega))$.*
(4) *We have $f\left(T(q), \omega_{\mathrm{can}}\right) \in A[[q]]$.*

This definition can also be extended to include level structure in the obvious way. If $f$ is a modular form over a ring $R$ which admits an injection $\phi : R \to \mathbf{C}$, then the image of the $q$-expansion of $f$ under $\phi$ coincides exactly with the usual definition of the $q$-expansion of a modular form in terms of its Fourier expansion.

1.3. **The $q$-expansion priniciple.** We have (Prop. 1.6 of [Kat73]):

1.3.1. **Proposition.** *A modular form $f$ is determined by its $q$-expansion.*

If one is to be precise, this applies only to *connected* modular curves $X(\Gamma)$; for non-connected curves one must (clearly) have the data of $f$ at a cusp on each component. Note that the basic idea of Prop. 1.3.1 is obvious — if $f$ vanishes on a Tate curve then it should vanish in a "neighbourhood" of infinity, and thus over the entire curve. In fact, this is exactly the proof, more or less. We also refer to the following corollary as the $q$-expansion principle:

1.3.2. **Corollary.** *Let $R \to S$ be an inclusion of rings, and suppose that the level $N$ of $\Gamma$ is invertible in $R$. Suppose that $f$ is a modular form over $S$ whose $q$-expansion has coefficients in $R[[q]] \subset S[[q]]$. Then $f$ arises from a modular form in $R$.*

This is an easy consequence of the previous lemma. What this theorem really means is that we can be a little sloppy with defining the rings $R$ we are working over, because we can "detect" the smallest such $R$ from the $q$-expansion.

1.4. **Hecke operators.** One often sees the Hecke operator $T_p$ on modular forms of weight $k$ to be defined as follows:

$$T_p\left(\sum a_n q^n\right) = \sum (a_{np} + p^{k-1} a_{n/p}) q^n,$$

where $a_{n/p}$ is interpreted to be zero unless $p$ divides $n$. While this is an important property of $T_p$ (which indeed characterizes it, by the $q$-expansion principle), it is a rubbish definition. Here are two better ones.

1.4.1. *$T_p$ as a correspondence.* Suppose that $R$ is a ring in which $p$ is invertible. We have a diagram as follows:

$$\begin{array}{ccc} X_0(p) & \xrightarrow[w_p]{} & X_0(p) \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ X & \xrightarrow[C_p]{} & X \end{array}$$

Recall that $X_0(p)$ is the moduli space of pairs $(E, D)$ together with a cyclic isogeny $\phi : E \to D$ of order $p$. The map $w_p$ sends a pair $\phi : E \to D$ to $\widehat{\phi} : D \to E$, where $\widehat{\phi}$ is the dual isogeny. It is an involution. The map $\pi$ is the natural projection. The

"map" $C_p$ is a many to many map which makes the diagram commute. Given a finite map $\phi : Z \to Z$ between varieties, the graph of $\phi$ in $Z \times Z$ defines a closed subscheme of (co-)dimension $\dim(Z)$. The map $C_p$ gives rise to a correspondence, which defines a Zariski closed subscheme of $X \times X$. Any such correspondences induces *actual* maps on any object functorially associated to $X$ which is *linear* — for example, the ring of functions on $X$, the tangent space of $\mathrm{Jac}(X)$, and more generally any appropriate cohomology group (coherent, Betti, étale) associated to $X$. The induced map $C_p$ then acts as $\pi_*(\pi \circ w_p)^* = \pi_* w_p^* \pi^*$, and in particular defines a map:

$$pT_p = \pi_*(\pi \circ w_p)^* : H^0(X(\Gamma), \omega^k) \to H^0(X(\Gamma), \omega^k).$$

Since $p$ is invertible in $R$, we may divide by $p$ to obtain $R$.

1.4.2. **Remark.** In the above calculation, we secretly made an identification $\pi^*\omega^k \simeq (\pi \circ w_p)^*\omega^k$. One way to do this is simply from the definition of $\omega$. On other natural Hecke modules (like $H^1(X, \mathbf{Q}_l)$) the correspondence $C_p$ induces $T_p$, rather than $pT_p$. The identification of $T_p$ as above with the map on $q$-expansions is done (for example) in [Buz03]. (We also do a related computation below.)

1.4.3. *$T_p$ on modular forms defined as a rule.* The definition of $T_p$ above can be made very explicit when modular forms are thought as sections of $H^0(X(\Gamma), \omega^k)$.

Let $(E, \alpha)$ denote a point on $X(\Gamma)$, where we denote by $\alpha$ the auxiliary level structure associated to the $E$. We assume that the level of $\Gamma$ is prime to $p$. The pre-image of the map:

$$X(\Gamma) \times_{X(1)} X_0(p) \to X(\Gamma)$$

consists of $p + 1$ points $\phi : D \to E$ where $\phi$ is cyclic of degree $p$. In any such situation, if $\omega \in H^0(E, \Omega)$, we may pull $\omega$ back to $D$ via $\phi^*$. Similarly, one may pull back $\alpha$ to a level structure $\phi^*\alpha$ on $D$.

1.4.4. **Definition.** *If $f$ is a modular form of weight $k$, and let $E/R$ be an elliptic curve where $p \in R$ is invertible. Then*

$$T_p f(E, \omega, \alpha) = p^{k-1} \sum_{\phi : D \to E} f(D, \phi^*(\omega), \phi^*\alpha).$$

Let us also introduce the operator $U_p$ which makes sense at level $\Gamma_0(p)$. Here we are given $E$ together with a distinguished $p$-isogeny $\eta : E \to B$. We define $U_p$ simply by considering the maps $\phi : D \to E$ which are *not* equal to $\eta^\vee : B \to E$. that is;

$$U_p f(E, \eta : E \to B, \omega, \alpha) = p^{k-1} \sum_{\phi : D \to E}^{\phi \neq \eta^\vee} f(D, \phi^*(\omega), \phi^*\alpha).$$

The equivalence of these definitions with the usual ones involving $q$-expansions is an easy exercise, which we now do. As usual, we ignore the level structure $\alpha$, because it doesn't make any difference to the computation.

One can also consider the same operator but now thinking of subgroup schemes $P \subset E[p]$ instead of maps $D \to E$, that is:

1.4.5. **Alternate Definition.** *If $f$ is a modular form of weight $k$, and let $E/R$ be an elliptic curve where $p \in R$ is invertible. Then*

$$T_p f(E, \omega, \alpha) = p^{k-1} \sum_{\phi: E/P \to E} f(E/P, \phi^*(\omega), \phi^*\alpha),$$

*where the sum is over all $p+1$ étale subgroup schemes $P$ of order $p$ in $E[p]$.*

1.4.6. **Exercise.** *Understand why — as promised in §1.3 — the above definition is sloppy. Hint: why is $T_p f$ a modular form over $R$ if the maps $\phi$ (or the subgroup schemes $P$) are not necessarily defined over $R$. Show that everything is OK using the $q$-expansion principle (Corr. 1.3.2).*

Given $T(q)$, we would like to write down the $p+1$ curves $D$ together with the corresponding isogenies $\phi_i : D \to T(q)$. For any such map, we have corresponding dual isogenies $\phi_i^\vee : T(q) \to D$, which are determined by the cyclic subgroup scheme of order $p$. The subgroups of order $p$ are given by the subgroups of the $p$-torsion, which is:

$$T(q)[p] = \{\zeta_p, q^{1/p}\}.$$

Hence we have the $p+1$ maps $\phi_0^\vee, \ldots, \phi_p^\vee$ defined as follows:

$$\phi_i^\vee := \begin{cases} T(q) \to T(q)/q^{1/p}\zeta_p^i, & i = 0, 1, \ldots, p-1 \\ T(q) \to T(q)/\zeta_p, & i = p. \end{cases}$$

The elliptic curve $T(q)/q^{1/p}\zeta^i$ is isomorphic to $T(q^{1/p}\zeta^i)$. On the other hand, the elliptic curve $T(q)/\zeta_p$ is isomorphic to $T(q^p)$ via the map induced by the $p$-th power map on $\mathbb{G}_m$. We may thus write down the corresponding dual isogenies as follows:

$$\phi_i = \begin{cases} T(q^{1/p}\zeta_p^i) \to T(q^{1/p}\zeta_p^i)/\zeta_p \simeq T(q), & i = 0, 1, \ldots, p-1 \\ T(q^p) \to T(q^p)/q = T(q), & i = p. \end{cases}$$

Let us suppose that

$$f\left(T(z), \omega_{\text{can}}\right) = \sum a_n z^n.$$

Then formally we have:

$$\begin{aligned} f\left(T(q), \omega_{\text{can}}\right) &= \sum a_n q^n, \\ f\left(T(q^{1/p}\zeta_p^i), \omega_{\text{can}}\right) &= \sum a_n q^{n/p}\zeta_p^{ni}, \\ f\left(T(q^p), \omega_{\text{can}}\right) &= \sum a_n q^{np}. \end{aligned}$$

The isomorphism $T(q^p)/q \simeq T(q)$ sits inside the commutative diagram:

$$\begin{array}{ccc} \mathbb{G}_m & = & \mathbb{G}_m \\ \downarrow & & \downarrow \\ T(q^p) & \xrightarrow[\phi_p]{} & T(q) \end{array}$$

and hence $\phi_p^*\omega_{\mathrm{can}} = \phi_p^* dt/t = dt/t = \omega_{\mathrm{can}}$. Hence, by definition:

$$\begin{aligned} T_p f(T(q), \omega_{\mathrm{can}}) &= p^{k-1}\left( f(T(q^p), \omega_{\mathrm{can}}) + \sum_{i=0}^{p-1} f(T(q^{1/p}\zeta_p^i), p\omega_{\mathrm{can}}) \right) \\ &= p^{k-1}\left( f(T(q^p), \omega_{\mathrm{can}}) + p^{-k}\sum_{i=0}^{p-1} f(T(q^{1/p}\zeta_p^i), \omega_{\mathrm{can}}) \right) \\ &= p^{k-1}\sum a_n q^{np} + \frac{1}{p}\sum_{i=0}^{p-1} a_n \zeta_p^{in} q^n \\ &= \sum_{n=0}^{\infty} (a_{np} + p^{k-1}a_{n/p}) q^n. \end{aligned}$$

recovering the previous definition.

1.4.7. *Hecke operators defined on functions of lattices.* Yes, you can do that too, if you like.

1.4.8. *The operator $T_p$ on $q$-expansions in characteristic $p$.* The definitions presented above included the assumption that $p$ be invertible in $R$. Yet the effect on $q$-expansions does not introduce denominators, and hence one may expect that the operator also exists in characteristic $p$ at level prime to $p$. That one can do this is an immediate consequence of the $q$-expansion principle, namely, one may lift to characteristic zero, apply $T_p$, and then reduce modulo $p$ to get the following commutative diagram defining $T_p$ in characteristic $p$:

$$\begin{array}{ccccc} H^0(X/\mathbf{Z}_p, \omega^k) & \xrightarrow{T_p} & H^0(X/\mathbf{Z}_p, \omega^k) & \hookrightarrow & \mathbf{Z}_p[[q]] \\ \downarrow & & \downarrow & & \downarrow \\ H^0(X/\mathbf{F}_p, \omega^k) & \xrightarrow{T_p} & H^0(X/\mathbf{F}_p, \omega^k) & \hookrightarrow & \mathbf{F}_p[[q]] \end{array}$$

1.4.9. *The operator $T_p$ on $q$-expansions in characteristic $p$ and weight one.* Perhaps you might complain that the argument above assumes that the mod-$p$ reduction map is surjective in weight one. The formula, however, is still correct. The point is that, after removing the cusps, $Y$ is affine, and then we have maps:

$$\begin{array}{ccc} H^0(Y/\mathbf{Z}_p, \omega^k) & \xrightarrow{T_p} & H^0(Y/\mathbf{Z}_p, \omega^k) \\ \downarrow & & \downarrow \\ H^0(Y/\mathbf{F}_p, \omega^k) & \xrightarrow{T_p} & H^0(Y/\mathbf{F}_p, \omega^k) \end{array}$$

If $\overline{f} \in H^0(X/\mathbf{F}_p, \omega^k)$, the lifted form $f$ may have poles at the cusps, but the definition of $T_p$ (as well as the computation involving $q$-expansions) still makes sense for meromorphic forms. The reduction

$$\overline{T_p(f)}$$

is then *a priori* a meromorphic modular form, however, by looking at the $q$-expansion, we see that it is regular at the cusps and thus holomorphic by the $q$-expansion principle. (This nice argument is due to Gross [Gro90] §4.) On the other hand, for discussion of Hecke operators "without the crutch of $q$-expansions" see [Con07].

1.5. **The Frobenius morphism.** Suppose that $S$ is a ring with $pS = 0$, and suppose that $X/S$ is a scheme. In this context, there are a pair of maps which go via the name Frobenius. First, one has the *absolute* Frobenius, which induces maps $F_{\text{abs}} : \text{Spec}(S) \to \text{Spec}(S)$ and $X \to X$. This map is given, locally on rings, by the map $x \mapsto x^p$. There is a commutative diagram as follows:

$$\begin{array}{ccc} X & \xrightarrow[F^*_{\text{abs}}]{} & X \\ \downarrow & & \downarrow \\ \text{Spec}(S) & \xrightarrow[F^*_{\text{abs}}]{} & \text{Spec}(S) \end{array}$$

In particular, $F^*_{\text{abs}}$ is not a map of schemes over $S$, unless $F^*_{\text{abs}}$ on $S$ happens to be constant (so $S = \mathbf{F}_p$). The relative Frobenius is a way of modifying this to give a morphism of schemes over $S$. Namely, let $X^{(p)} = X \times_S S$, where $S$ is thought of over $S$ not by the trivial map but via $F_{\text{abs}}$. Then, by construction, there is a map, *relative* Frobenius, given by $F : X \to X^{(p)}$, such that the composition with the natural map $X^{(p)} \to X$ is $F^*_{\text{abs}}$.

1.5.1. **Exercise.** *Let $X$ be the smooth curve:*

$$ax^3 + by^3 + cz^3 = 0$$

*over $k$, where $3abc \neq 0$. Prove that $X^{(p)}$ is the curve*

$$a^p x^3 + b^p y^3 + c^p z^3 = 0.$$

Note that the map $F^*_{\text{abs}}$ on $H^*(X, \mathcal{O}_X)$ is *not* $S$-linear, although it is $\mathbf{F}_p$-linear.

1.6. **The Hasse invariant.** Let $S$ be as in the previous section. Suppose that $E/S$ is an elliptic curve together with a differential $\omega_S$ generating $\Omega^1_{E/S}$. By Serre duality, we may associate to $\omega_S$ a dual basis element $\eta \in H^1(E, \mathcal{O}_E)$. The Frobenius map induces a map:

$$F^*_{\text{abs}} H^1(E, \mathcal{O}_E) \to H^1(E, \mathcal{O}_E),$$

and we may write

$$F^*_{\text{abs}}(\eta) = A(E, \omega) \cdot \eta$$

for some $A(E, \omega) \in S$.

1.6.1. **Lemma.** *$A$ is a meromorphic modular form of level one and weight $p - 1$ over $S$.*

*Proof.* If one replaces $\omega$ by $\omega' = \lambda \cdot \omega$ for $\lambda \in S^\times$, then $\eta$ becomes $\eta' = \lambda^{-1}\eta$, and

$$F^*_{\text{abs}}(\eta') = F^*_{\text{abs}}(\lambda^{-1}\eta) = \lambda^{-p} A(E, \omega) \cdot \eta = \lambda^{1-p} A(E, \omega) \cdot \eta',$$

and thus

$$A(E, \lambda \cdot \omega) = \lambda^{1-p} A(E, \omega).$$

□

If $S$ is a field, then $A(E,\omega)$ is either zero or a unit — it is zero precisely when $E$ is supersingular (this is essentially a definition of what it means to be supersingular, see [Sil86]). We call $A$ the *Hasse invariant* of $E$. To compute the $q$-expansion of $A$, we need to evaluate $A$ on the pair $(T(q),\omega_{\mathrm{can}})$. In order to do this, we need to understand the operator $F^*_{\mathrm{abs}}$ more explicitly on curves.

1.7. **The Cartier operator on curves.** Suppose that $X/S$ is a smooth curve of genus $g$, and that $S$ is perfect. The *Cartier operator* $\mathcal{C}$ defines a map from the meromorphic differentials on $X$ to itself satisfying the following properties:

(1) $\mathcal{C}$ preserves the holomorphic differentials $H^0(X/S,\Omega^1)$.
(2) $\mathcal{C}(f^p\omega) = f\cdot\mathcal{C}(\omega)$ for any meromorphic function $f$.
(3) $\mathcal{C}(f^{n-1}df) = \begin{cases} df, & n=p \\ 0, & 0<n<p. \end{cases}$.
(4) If $\eta\in H^1(X,\mathcal{O}_X)$, then, under the pairing of Serre duality,
$$\langle\mathcal{C}(\omega),\eta\rangle = \langle\omega, F^*_{\mathrm{abs}}\eta\rangle.$$

To imagine why such an operator might exist, consider the completion $\widehat{\mathcal{O}_{X,x}}$ of a local ring at $x$. Because $X$ is smooth, the corresponding ring is $S[[x]]$, the meromorphic differentials are of the form $S[[x]]dx$. Then $\mathcal{C}$ may be defined as follows, writing $\omega = f(x)dx$:
$$\mathcal{C}(\omega) := \sqrt[p]{-\frac{d^{p-1}f}{dx^{p-1}}}dx.$$
Clearly $\mathcal{C}$ preserves holomorphicity at a point. By the chain rule, one has:
$$\frac{d^{p-1}f^pg}{dx^{p-1}} = f^p\frac{d^{p-1}g}{dx^{p-1}},$$
$$\frac{d^{p-1}f^{n-1}df}{dx^{p-1}} \approx \frac{1}{n}\cdot\frac{d^p}{dx^p}(f^n).$$
The usage of $\approx$ is meant to indicate that this equality is only an equality of formal expressions over $\mathbf{Z}[[x]]$ — over $S[[x]]$ it makes perfect sense if $(n,p)=1$ and implies that $\mathcal{C}(f^{n-1}df)=0$, because $d^p/dx^p$ is clearly the zero operator. For $p|n$ it is not too hard to make formal sense of what this means and compute that $\mathcal{C}(f^{p-1}df)=df$. To make this rigorous, one needs to show that $\mathcal{C}$ may also be defined algebraically, and that it does not depend on the choice of a uniformizer. It would take us a little too far afield to prove these statements, however.

1.7.1. *Deligne's Computation of $A$.* Now let us compute the $q$-expansion of $A$, namely, to compute $A(T(q),\omega_{\mathrm{can}})$ over $R=\mathbf{F}_p((q))$.

1.7.2. **Theorem** (Deligne)**.** $A(T(q),\omega_{\mathrm{can}})=1$.

The idea is simply to show that the corresponding 1-form $\eta$ is preserved under $F^*_{\mathrm{abs}}$. From the characterizing properties of the Cartier operator, it thus suffices to show that $\mathcal{C}(\omega_{\mathrm{can}})=\omega_{\mathrm{can}}$. Let $x=t-1$ be a uniformizing parameter of $\mathbb{G}_m$ at the origin. Then a ($\mathbb{G}_m$-invariant) differential is given by
$$\frac{dx}{1+x} = \frac{dt}{t} = \omega_{\mathrm{can}}.$$
Yet an easy computation shows that $\mathcal{C}(dt/t) = t^{-1}\mathcal{C}(t^{p-1}dt) = dt/t$. We can't quite argue this way, because $R$ is not perfect (even the local definition of $\mathcal{C}$ above involves taking $p$th roots and thus requires that the underlying ring $S$ be perfect).

On the other hand, if $R^{\mathrm{per}}$ is the perfection of $R$, then, since $R \to R^{\mathrm{per}}$ is flat, by the $q$-expansion principle (Corr. 1.3.2), it suffices to work over $R^{\mathrm{per}}$, where the argument above goes through.

1.8. **Lifting the Hasse invariant.** (cf. [Buz03]). If $p \geq 5$, then $A$ lifts to a modular form in characteristic zero. From the computation above, we note that the $q$-expansion of any such lift is congruent to $1 \mod p$. From the Kummer congruences, we deduce:

1.8.1. **Theorem.** *Suppose that $p \geq 5$. The modular form $E_{p-1}$ is a lift of $A$ such that $E_{p-1} \equiv A \mod p$. If $p = 2$ or $p = 3$, the modular forms $E_4$ and $E_6$ are lifts of $A^4 \mod 8$ and $A^3 \mod 9$ respectively.*

*Proof.* For $p = 2$ and 3 this is a computation; for $p \geq 5$ it is an immediate consequence of the von Staudt–Clausen theorem on Bernoulli numbers, as well as the identification of the constant term of the classical holomorphic Eisenstein series. □

Just to be clear, here by "mod" $p$ we really mean modulo the ideal $(p)$, so that if $E_{p-1}$ is thought of as a modular form over some ring $R$, this congruence identifies the value of $A$ in $R/p$, even if the latter is not reduced. In particular, given a Weierstrass equation for $E$ (and hence a canonical differential $\omega$) one may compute $A \mod p$ by computing the corresponding value of $E_{p-1}$, with appropriate modifications if $p = 2$ or 3.

## 2. $p$-adic modular forms

Let us fix a congruence subgroup $\Gamma$ of level prime to $p$. The definition of $p$-adic modular form and overconvergent $p$-adic modular form at level one are virtually the same as the corresponding definition at level $\Gamma$ — one need only add the natural compatibility with the level structure $\alpha$ away from $p$. In the sequel, therefore, we shall essentially ignore this distinction and work at level one, making remarks about the level structure away from $p$ (the "tame" level structure) when appropriate.

2.1. **$p$-adic modular forms: The Serre approach.** Serre wrote a beautiful[3] and elementary paper on $p$-adic modular forms [Ser73b]. The basic idea (translated in to somewhat different language) is as follows. In order to capture the notion of congruences between modular forms in some topological way, then we would like to say that two $q$-expanions $a$ and $b$ are *close* if $a \equiv b \mod p^n$ for large $n$. Recall that (up to normalization) the space of modular forms has a basis with coefficients in $\mathbf{Z}$. There is a natural topology on $\mathbf{Z}_p[[q]] \otimes \mathbf{Q}_p$ (note, this is different from $\mathbf{Q}_p[[q]]$)

[3]As one would expect, Serre effortlessly explains everything seemingly starting from first principles and gives a beautiful explanation of the construction of $p$-adic $L$-functions. Following the elementary arguments down to the source, the key fact is to show that $\phi = \sum_{n=1}^{\infty} \sigma_{p-2}(n) q^n$ does not lie in the field of fractions of the ring of modular forms modulo $p$ on the complement $X \setminus S$ of the supersingular locus ([Ser73b], p.199 Ser-9). An elementary argument using weights shows that $\phi$ itself is not a mod-$p$ modular form. Serre notes that $\phi - \phi^p = \psi$ for some explicit $\psi$, and then uses the fact that $H^0(X \setminus S, \mathcal{O}_X)$ is *integrally closed* (because $X$ and thus $X \setminus S$ is smooth) to obtain a contradiction. Yet to get the identification of $\widetilde{M^0}$ with $H^0(X \setminus S, \mathcal{O}_X)$, one uses the fact that the Hasse invariant $A$ is congruent to $1 \mod p$, which is of an order of difficulty higher than the rest of the arguments in the paper. Thus it is better to read [Ser73b] in conjuction with Serre's Bourbaki seminar on the subject [Ser73a], which gives a little more detail concerning this argument.

which exactly records this notion of congruence, and we may define $p$-adic modular forms to be the closure of the set of modular forms (see [Ser73b]). Let $A$ be (any) lift of the Hasse invariant. Since $A \equiv 1 \mod p$, it follows that the powers of $A$ (at least the $p^n$th powers) are becoming more and more congruent to 1 mod $p$. Hence they converge to 1. It follows that the powers $A^{p^n-1}$ are converging to $A^{-1}$, or in particular that *any lift of the Hasse invariant is invertible.* The space of $p$-adic modular functions then defines itself:

2.1.1. **Definition.** *The $p$-adic modular functions on $X(\Gamma)$ are the functions which are well defined at all points of ordinary reduction.*

Naturally enough, one might be a little suspicious of this definition, since one is allowing poles at an infinite number of supersingular points (there are only finitely many supersingular points modulo $p$, but there are infinitely many lifts to characteristic zero). An initial step to repairing this is give a rule–based definition.

2.1.2. **Definition** (Version 2)**.** *A $p$-adic modular form $f$ of weight $k$ and level one over a $p$-adically complete algebra $A$ is a function on pairs $(E/R, \omega)$ for a $p$-adically complete $A$-algebra $R$ satisfying*

(1) *$\omega$ is a nowhere vanishing section of $\Omega^1_{E/A}$,*
(2) *$A(E/B, \omega_B)$ is invertible, where $B = A/p$,*

*such that:*

(1) *$f(E/A, \omega)$ depends only on the $A$-isomorphism class of $(E/A, \omega)$.*
(2) *$f(E, \mu\omega) = \mu^{-k} f(E, \omega)$ for any $\mu \in A^\times$.*
(3) *If $\phi : A \to B$ is any map of rings, then $f(E/B, \omega_B) = \phi(f(E/A, \omega))$.*
(4) *$f(T(q), \omega_{\mathrm{can}}) \in A[[q]]$.*

As expected, there is an analogous definition at level $\Gamma$ prime to $p$, where one considers functions $f(E/A, \omega, \alpha)$ for some level structure $\alpha$ away from $p$ corresponding to the group $\Gamma$. Denote this space by $M_k(\Gamma, R, 0)$. If $\Gamma = \mathrm{SL}_2(\mathbf{Z})$ (which is a natural choice for which all the phenomena can already be seen) we simply write $M_k(1, R, 0)$. Clearly any classical modular form of weight $k$ over $R$ gives a $p$-adic modular form. Moreover, a $p$-adic modular form over a finite field $k$ (necessarily of characteristic $p$) consists of sections of $H^0(X \setminus S, \omega^k)$, where $S$ is the supersingular locus (since invertible over a field is the same as non-vanishing). If $A$ is the Hasse invariant, then $A$ is a modular form of weight $p-1$ over $\mathbf{F}_p$, and $A^{-1}$ is a modular form of weight $1-p$ over $\mathbf{F}_p$. Moreover, $A^{p^{n-1}}$ defines an invertible $p$-adic modular form of weight $p^{n-1}(p-1)$ over $\mathbf{Z}/p^n\mathbf{Z}$, and thus:

2.1.3. **Lemma.** *Suppose that $p^n = 0$ in $R$. Then the map;*

$$A^{p^{n-1}} : M_k(\Gamma, R, 0) \to M_{k+p^{n-1}(p-1)}(\Gamma, R, 0)$$

*is an isomorphism.*

From this we deduce the $q$-expansion principle.

2.1.4. **Lemma.** *There is an injective map $M_k(\Gamma, R, 0) \to R[[q]]$.*

*Proof.* By construction,

$$M_k(\Gamma, R, 0) = M_k(\Gamma, \varprojlim R/p^n, 0) = \varprojlim M_k(\Gamma, R/p^n, 0).$$

It thus suffices to assume that $p^n = 0$ in $R$. By dévissage (and Prop. 1.2.24), it suffices to consider the case $R = \mathbf{Z}/p^n\mathbf{Z}$. Given $g \in M_k(\Gamma, R, 0)$, its reduction modulo $p$ extends to a section of $H^0(X, \omega^k)$ with poles of finite orders at the supersingular points. In particular, after multiplication by some power of $A^{p^{n-1}}$, it extends to a classical modular form. Since the $q$-expansion of $A^{p^{n-1}}$ is 1, this construction does not depend on any choices. The lemma then follows from the classical $q$-expansion principle. □

Moreover:

2.1.5. **Lemma.** *The closure of the set of classical modular forms over a $p$-adically complete ring $R$ of all weights coincides with the set of $p$-adic modular forms over $R$ of all weights.*

*Proof.* Again by the limit property of $M_k(\Gamma, R, 0)$ mentioned above, It suffices to show that any classical modular form $f$ gives an element of $M_k(\Gamma, R/p^n, 0)$ where

$$\text{weight}(f) \equiv k \mod p^{n-1}(p-1),$$

and conversely that every element of the latter set arises in such a way. As in the previous lemma, any classical $f$ defines a modular form in $M_{\text{weight}(f)}(\Gamma, R/p^n, 0)$ which then only depends on $\text{weight}(f) \mod p^{n-1}(p-1)$ by Lemma 2.1.3. Conversely, the construction of $q$-expansions above implies that any $g$ comes from a space of classical modular forms modulo $p^n$ of large weight, which then lifts to a classical modular form by Lemma 1.2.24. □

Another nice property of $p$-adic modular forms is that one sees all forms of $p$-power level:

2.1.6. **Lemma.** *Suppose that $R$ is a $p$-adically complete ring such that $\zeta_{p^n} \in R$. Then there is an inclusion:*

$$M_k(\Gamma_1(p^n), R) \subset M_k(1, R, 0)$$

*for any $k$.*

*Proof.* By dévissage (and Prop. 1.2.24) it suffices to consider the case $R = \mathbf{Z}_p[\zeta_{p^n}]$. Let $f$ be a classical modular form of level $\Gamma_1(p^n)$. Suppose we are given an elliptic curve $(E/R, \omega)$ such that $A(E_S, \omega_S)$ is a unit for $S = R/\pi$ so $E$ is ordinary. The group scheme $E[p^n]$ is then an extension of an étale group scheme by a local group scheme $C$. The base change of $C$ to $S = R/p$ is the kernel of $\text{Frob}_p^n$, and (for example) if $R$ is the ring of integers of a finite extension of $\mathbf{Q}_p$ then the $\overline{\mathbf{Q}}_p$-points of $C$ are the kernel of the reduction map. Over some unramified extension of $R$, there is an isomorphism $C \simeq \mu_{p^n}$. Since $\zeta_{p^n} \in R$, this module has a canonical generator $P$, and hence we define the $p$-adic modular form $g$ by

$$g(E, \omega) = f(E, \omega, P).$$

□

A similar argument shows that $M_k(\Gamma_0(p^n), R) \subset M_k(1, R, 0)$ for any $p$-adically complete ring $R$.

2.1.7. **Exercise.** *Serre* [Ser73b] *also proves this result using the level lowering properties of the $U_p$ operator. Show that these proofs are the same.*

2.1.8. *Congruences for p-adic modular forms.* Given a $p$-adic modular form $f$ over $R$, we have the following basic result governing congruences for $f$:

2.1.9. **Lemma.** *Let $R$ denote the ring of integers of a finite extension of $\mathbf{Q}_p$. Let $f \in M_k(\Gamma, R, 0)$, where $\Gamma$ has level $N$. For any fixed power $p^d$ of $p$, the quantity*

$$a(\ell) \mod p^d$$

*depends only on the conjugacy class of $\mathrm{Frob}_\ell$ in a finite extension $K_d/\mathbf{Q}$ unramified outside $Np \cdot \infty$.*

*Proof.* By passing to sufficiently high integral weight, we may lift $f \mod p^n$ to a classical modular form $g$. We may then write $g$ as a finite sum of eigenforms, where $\alpha_i \in \overline{\mathbf{Q}}_p$:

$$g = \sum \alpha_i g_i.$$

In the usual manner [Del71], the eigenforms $g_i$ may be associated with Galois representations unramified outside $Np \cdot \infty$, and thus the coefficient $a_i(\ell)$ of $g_i$ mod any fixed power of $p$ only depends on the conjugacy class of $\mathrm{Frob}_\ell$ in some finite extension unramified outside $Np \cdot \infty$. The result follows. □

Note that as $n$ increases, the number of eigenforms $g_i$ required typically increases, and the fields $K_d$ become more and more complicated. For certain exceptional $g$ and for small $d$, however, the fields $K_d$ may turn out to be abelian, in which case $a(\ell)$ only depends on $\ell$ modulo some fixed modulus (by class field theory).

2.1.10. **Exercise.** *Let $g = \sum a(n)q^n$ be a $p$-adic modular form. For any integer $d$, prove that there exists a positive density of primes $\ell$ such that $a(\ell) \equiv 0 \mod p^d$. Hint: use the Cebotarev density theorem and the fact that $\mathrm{Tr}(\rho(c)) = 0$ for a modular Galois representation $\rho$ and any complex conjugation $c \in G_{\mathbf{Q}}$.*

2.1.11. *Rigid Analytic Spaces.* Let $X$ be a modular curve which is smooth over $\mathbf{Z}[1/p]$. It makes perfect sense to talk about the ordinary locus of $X/\mathbf{F}_p$, since there are only finitely many supersingular points. It makes less sense to talk about the "ordinary locus" over $\mathbf{Z}_p$, since know we would like to exclude all (infinitely many) lifts of supersingular elliptic curves. Specifically, we would like to remove a "unit ball" around any supersingular point. Clearly this is not a construction that can be done in the Zariski topology. Rigid analytic spaces provide the right context in which these constructions make sense, and such that the topology is fine enough to allow such constructions.

2.1.12. **Exercise.** *Pick up a book on rigid analytic spaces. Hold it in your hand held out perpendicularly from your body for approximately ten minutes. Deduce that actually reading it would be less painful and stop complaining already.*

Alternatively, consult Brian Conrad's lectures from the 2007 Arizona Winter School [BCD+08].

For a modular curve $X$, a first approximation to thinking about the associated rigid analytic space $X^{\mathrm{rig}}$ is to think about $X$ as a complex manifold over $\mathbf{C}_p$.

2.1.13. **Definition.** *The $p$-adic modular forms of weight $k$ are the global sections $H^0(X^{\mathrm{rig}}[0], \omega^k)$, where $X^{\mathrm{rig}}[0]$ denotes the ordinary locus of the rigid analytic space $X^{\mathrm{rig}}$.*

Instead of defining rigid analytic spaces, we consider an example which tells you everything you need to know (at least for the purposes of these lectures). Suppose that $N = 1$ and $p = 2$. The modular curve $X(1)$ is just the $j$-line. A curve $E/\mathbf{F}_2$ is supersingular if and only if $j_E = 0$. Thus the supersingular locus is the region $|j| < 1$, and the ordinary locus is the region $|j| \geq 1$, or $|j^{-1}| \leq 1$. In particular, the $p$-adic modular functions of level one for $p = 2$ are exactly given by the Tate algebra:

$$\mathbf{C}_2\langle j^{-1}\rangle := \sum_{n=0}^{\infty} a_n j^{-n}, \qquad \lim |a_n| = 0.$$

That is, the 2-adic modular functions of level one are just functions on an explicit closed ball, and $X^{\mathrm{rig}}[0]$ consists of the maximal ideals of this ring, which are easily seen to consist canonically of points in this ball.

2.2. **The ordinary projection.** Let $R$ be $p$-adically complete ring, and consider the space $M_k(\Gamma_0(p), R)$ of classical modular forms. The operator $U_p$ acts on this space. Since this space is finite as an $R$-module, we may define the following operator:

$$e_p := \lim_{\to} U_p^{n!}.$$

2.2.1. **Lemma.** *$e_p$ is an idempotent on $M_k(\Gamma_0(p), R)$, and projects onto the space generated by Hecke eigenforms on which $U_p$ acts by a unit.*

*Proof.* By dévissage (and Prop. 1.2.24), we may reduce to the case when $R = \mathbf{Z}_p$. In this case, the result is an elementary statement in linear algebra. □

Note that $U_p$ and hence $e_p$ commutes with the Hecke operators $T_\ell$ for $\ell$ prime to the level. Thus the image (and kernel) of $e_p$ is module for the Hecke algebra, and $e_p$ is an Hecke equivariant projection. If $f$ is a Hecke eigenform with unit eigenvalue for $U_p$ then one says that $f$ is *ordinary*.

The following is a consequence of results of Hida:

2.2.2. **Theorem** (Hida)**.** *The operator $e_p$ extends to an idempotent on $M_k(\Gamma, R, 0)$. Denote the image of $e_p$ by $e_p M_k(\Gamma, R, 0)$. Then*

1. *$e_p M_k(\Gamma, R, 0)$ is finite dimensional, and the dimension depends only on $k \bmod p - 1$.*
2. *If $k > 1$, then $e_p M_k(\Gamma, R, 0) \subseteq M_k(\Gamma_0(p), R, 0)$ is spanned by classical modular forms.*
3. *The minimal polynomial of $U_p$ on $e_p M_k(\Gamma, R/p^n, 0)$ only depends on $k \bmod p^{n-1}(p-1)$.*

In fact, Hida proves much more than this. The last condition points to the fact that the eigenforms of weights $k$ and $k'$ are congruent to each other modulo some power of $p$ which depends on $(k - k')$ — in fact one can form a natural *family* $\mathcal{H}$ of ordinary modular eigenforms that varies over all weights, in particular a finite module $\mathcal{H}$ over $\Lambda = \mathbf{Z}_p[[\mathbf{Z}_p^\times]]$ such that

$$\mathcal{H} \otimes_\Lambda \mathbf{Z}_p(\psi_k) \simeq e_p M_k(\Gamma_0(p), \mathbf{Z}_p),$$

where $\mathbf{Z}_p(\psi_k)$ is the abelian group $\mathbf{Z}_p$ where the action of $\mathbf{Z}_p^\times$ is via the character $\psi_k(x)m = x^{k-1}m$. Indeed the theme of eigenforms *varying in families* is central to the topic of $p$-adic and overconvergent modular forms, although we concentrate on somewhat different topics in these notes.

2.2.3. **Remark.** *Since newforms of level* $\Gamma_0(p)$ *have* $U_p$*-eigenvalue equal to one of the numbers* $\pm p^{(k-2)/2}$*, when* $k > 2$*, the ordinary projection is generated by old forms from* $M_k(\Gamma, R)$*.*

2.3. **Why $p$-adic modular forms are not good enough.** Let us explain why one might ask to go *beyond* the theory of $p$-adic modular forms. We do this with an example, first observed by Atkin and O'Brien [AO67]. The $j$-invariant defines a meromorphic modular form of weight zero and level 1 for any $p$. It follows that $U_p j$ also a *$p$-adic* meromorphic modular form of weight zero and level 1. On the other hand, since $U_p j \in \mathbf{Z}_p[[q]]$, the function $U_p j$ extends to the cusp and defines an honest $p$-adic modular form. By Hida's theorem, it follows that $e_p U_p j$ lies in the ordinary space $e_p M_0(1, \mathbf{Z}_p, 0)$, and can thus be written as a finite sum of (generalized) ordinary eigenforms. Atkin and O'Brien consider the special case of $p = 13$. In this case, the dimension of $e_{13} M_0(1, \mathbf{Z}_{13}, 0)$ is the same dimension as $e_{13} M_{12}(1, \mathbf{Z}_{13}, 0)$, which is the image of the classical space $M_{12}(1, \mathbf{Z}_{13})$ under $e_{13}$. The space of modular forms of weight 12 and level 1 over $\mathbf{Z}_{13}$ is generated by $E_{12} \equiv 1 \mod 13$ and $\Delta$. The eigenvalue of $T_{13}$ on $\Delta$ is

$$\tau(13) = -577738,$$

and thus the ordinary projection $e_{13} M_{12}(\Gamma_0(13), \mathbf{Z}_{13}, 0)$ consists of the forms:

$$E_{12}(q) - 13^{11} E_{12}(q^{13}), \qquad \Delta(q) - \beta \Delta(q^{13}),$$

with $U_{13}$ eigenvalues 1 and $\alpha$ respectively, where

$$\beta = -288869 - \sqrt{-1708715094876} = 5 \cdot 13^{11} + 3 \cdot 13^{12} + 9 \cdot 13^{13} + \dots$$
$$\alpha = -288869 + \sqrt{-1708715094876} = 8 + 5 \cdot 13 + 10 \cdot 13^3 + 5 \cdot 13^4 + \dots$$

are the roots (in $\mathbf{Z}_{13}$) of $x^2 + 577738x + 1792160394037 = 0$. In particular, the ordinary space has dimension two. *A priori* it is easy to obtain an upper bound of 2, but the lower bound requires the computation $\tau(13) \equiv 8 \not\equiv 0 \mod 13$. It follows that $e_{13} M_0(\Gamma, \mathbf{Z}_{13}, 0)$ has dimension two. The function 1 is a classical modular form of level one and weight 0, and it is also an ordinary form with eigenvalue 1. The other ordinary form is thus a normalized cuspidal eigenform $h \in \mathbf{Z}_p[[q]]$. Indeed, $h \equiv \Delta \mod 13$. It follows that there is an identity

$$e_{13} U_{13} j = 744 + \alpha h$$

for some $\alpha \in \mathbf{Z}_{13}$. In particular, we deduce that, for any fixed $d$,

$$U_{13}^n j - 744 \mod 13^d = \sum_{k=1}^{\infty} c(13^n k) q^k \mod 13^d$$

is a Hecke eigenform for sufficiently large $n$ depending on $d$. These arguments do not make clear, however, how large $n$ has to be for any particular $d$. On the other hand, Atkin and O'Brien conjectured something much more precise, namely, they conjecture that

$$\sum_{k=1}^{\infty} c(13^n k) q^k \mod 13^n$$

is already a Hecke eigenform. That is, the *convergence* of $U_{13}^n j$ to the ordinary projection is *linear*. We may ask: is this a general phenomenon for all $p$-adic modular forms? In this generality, it turns out that the answer is no.

2.3.1. *There are too many p-adic modular forms.* Consider the space of $p$-adic modular forms of level 1 and some weight, say $k = 12$. This certainly contains the modular form $\Delta$.

2.3.2. **Exercise.** *If* $f = \sum a_n q^n \in M_k(\Gamma, R, 0)$, *show that*

$$V_p f = \sum a_n q^{np} \in M_k(\Gamma, R, 0),$$

*and that* $U_p V_p$ *is the identity. Prove it by defining* $V_p$ *in the correct way. Then compare your nice argument to the explicit computations in* [Ser73b] *and smile to yourself.*

Let $g$ be the $p$-adic modular form

$$g := (1 - V_p U_p)\Delta = \sum_{(n,p)=1} \tau(n) q^n.$$

Note that $U_p g = (U_p - U_p V_p U_p)\Delta = (U_p - U_p)\Delta = 0$. Let $R = \mathcal{O}_{\mathbf{C}_p}$, which is $p$-adically complete. If $\lambda \in \mathbf{C}_p$ has positive valuation, then

$$f_\lambda := \sum_{n=0}^{\infty} \lambda^n V_p^n g$$

is also a $p$-adic modular form over $R$. On the other hand, since $U_p g = 0$ and $U_p V_p$ is the identity, one checks that $U_p f_\lambda = \lambda f_\lambda$. In particular, $f_\lambda$ is an eigenform for $U_p$ with eigenvalue $\lambda$. It is also easy to check that $f_\lambda$ is an eigenform for all the Hecke operators $T_\ell$ as well. Since $v(\lambda) > 0$, it follows that $e_p f_\lambda = 0$. But this implies that the ordinary projection might converge arbitrarily slowly in general, e.g., if one takes the $p$-adic modular form:

$$h = f_p + p f_{p^{1/2}} + p^2 f_{p^{1/3}} + \ldots = \sum f_{p^{1/n}} p^n \in \mathbf{C}_p[[q]].$$

Then:

(1) $e_p h = 0$,
(2) For arbitrarily small rational $r > 0$, we have $U_p^n h \not\equiv 0 \mod p^{nr}$ for sufficiently large $n$.

In particular, the convergence of $h$ is not linear (and one can cook up this example to make the convergence as slow as one desires).

In general, the fact that $U_p$ contains a continuous spectrum on $M_k(\Gamma, R, 0)$ rules out the possibility that one can decompose a $p$-adic modular function into an infinite sum of eigenforms. The key observation, already in [Kat73], is that one must consider sections of $X^{\mathrm{rig}}$ which converge *beyond* the ordinary locus $X^{\mathrm{rig}}[0]$. A key argument with $p$-adic modular forms is that one can pass between level 1 and level $\Gamma_0(p)$ (and higher levels) using the fact that an ordinary elliptic curve $E/R$ comes with a canonical subgroup scheme $P \subset E[p]$, coming from the kernel of the reduction map. The key idea turns out to be generalizing this construction to elliptic curves which are no longer ordinary.

## 3. The canonical subgroup

(cf. §3 of [Kat73].) Let $(R, \mathfrak{m})$ be the ring of integers of some finite extension of $\mathbf{Q}_p$ with residue field $R/\mathfrak{m} = k$ and fraction field $K$. Let us normalize valuations

so that $v(p) = 1$. If $E/R$ is *ordinary*, then $E(\overline{k})[p] = \mathbf{Z}/p\mathbf{Z}$. There is a reduction map:

$$E(\overline{K}) \to E(\overline{k}).$$

The kernel $C$ of $E(\overline{K})[p] \to E(\overline{k})[p]$ is thus a cyclic subgroup of order $p$, which is canonically associated to $E/R$.

Suppose now that $E/k$ is supersingular, and thus $E(\overline{k})[p]$ is trivial. The group $E(\overline{K})[p] = (\mathbf{Z}/p\mathbf{Z})^2$ contains $p+1$ subgroups $C$, and there does not seem to be any obvious way to make a canonical choice amongst all such subgroups. An idea of Lubin, however, shows that this *can* be done in many — although not all — cases. Let us do so explicitly when $p = 2$. An elliptic curve $E/R$ has a minimal Weierstrass equation of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

where $a_i \in R$. The choice of Weierstrass equation also determines a differential

$$\omega = \frac{dx}{2y + a_1x + a_3} \in H^0(E/R, \Omega^1).$$

Let $K$ be the field of fractions of $R$. Then we may explicitly find the 2-torsion points of $E$ over $\overline{K}$ in the usual way, in particular, we may write:

$$\left(y + \frac{a_1x}{2} + \frac{a_3}{2}\right)^2 = x^3 + \left(a_2 + \frac{a_1^2}{4}\right)x^2 + \left(a_4 + \frac{a_1a_3}{2}\right)x + \left(a_6 + \frac{a_3^2}{4}\right),$$

and then the $x$-coördinates[4] of the 2-torsion points are obtained by finding the roots of the cubic. How might one find a *canonical* root of this cubic? One way would be to look at the 2-adic valuation of the roots, and take the root of *minimal* valuation (if it exists). Note that the valuation of the roots can easily be computed by Newton's Lemma. Moreover, this naturally generalizes what happens in the ordinary case, where there will be exactly one root with valuation 0 which reduces to the 2-torsion point in $\overline{k}$. Let us now apply Newton's Lemma.

Suppose that $a_1 = 0 \mod \mathfrak{m}$. If $a_3 = 0 \mod \mathfrak{m}$, then the equation would be singular. Hence $v(a_3) = 0$. If $v(a_1) \geq 1$, then the valuation of the coefficients of the cubic are all non-negative except the last term, which will have valuation $-2$. In this case, all the roots have valuation $-2/3$, and there is no unique smallest root. Suppose instead that $v(a_1) < 1$. Then the coefficients have the following valuations:

$$[0, 2v(a_1) - 2, v(a_1) - 1, -2].$$

In particular, we have the following:

(1) If $v(a_1) \geq 2/3$, then all the roots have valuation $-2/3$.
(2) If $v(a_1) < 2/3$, then there is a unique root with valuation $2(v(a_1) - 1)$.

Strictly speaking, this calculation was only valid for $v(a_1) \neq 0$, but one can check that when $v(a_1) = 0$ that $E$ is ordinary, and that there is a unique root of valuation 0.

One might ask: what is the meaning of the coefficient $a_1$? The Weierstrass equation when written in this form is only well defined up to certain transformations. However, one can check that all such transformations leave $a_1 \mod 2$ invariant. In particular, $a_1 \mod 2$ is a well defined invariant of $E$ together with a choice of

[4] as the New Yorker would say.

differential, which is to say that $a_1$ is a modular form of level 1 and weight 1 over any ring $R$ in which $2 = 0$.

3.0.1. **Exercise.** *Let $S = R/2$. Prove that $a_1 \mod 2 = A(E_S, \omega_S)$ is the Hasse invariant.*

In particular, we have seen the following:

3.0.2. **Lemma.** *Let $R$ be a complete $\mathbf{Z}_2$-algebra, and $S = R/2$. An elliptic curve $(E/R, \omega)$ has a canonical subgroup of order 2 if and only if $v(A) < 2/3$, where $A = A(E_S, \omega_S)$ is the Hasse invariant of $E_S$.*

3.0.3. **Exercise.** *Consider the following curve*

$$E : y^2 + \sqrt{2} \cdot xy + y = x^3 - x$$

*over $R = \mathbf{Z}_2[\sqrt{2}]$. Prove that $E$ has a canonical subgroup generated over $K = \mathbf{Q}_2(\sqrt{2})$ by the point*

$$P = [2^{-1} + 2^{1/2} + 2^{3/2} + 2^2 + \ldots, 2^{-3/2} + 2^{-1} + 2^{-1/2} + 2^{1/2} + \ldots]$$

3.0.4. **Exercise.** *Do the same computation as was done in the beginning of this section except now with 3-torsion instead of 2-torsion. Show that there exists a canonical subgroup of order three if the valuation of $a_2$ is less than $3/4$. Identity $a_2 \mod 3$ with the Hasse invariant $A(E_S, \omega_S)$, where $S = R/3$.*

If $R$ is a discrete valuation ring and $E/R$ is an elliptic curve, then there is essentially a canonical choice of differential $\omega_R \in H^0(E/R, \Omega^1)$, because the latter is free of rank one over $R$ (canonical up to units in $R$, of course). It is elementary to see that the valuation of $A(E_S, \omega_S)$ with $S = R/p$ does not depend on this choice, so by abuse of notation we can talk of the valuation of $A(E_S)$. Suppose that $p \geq 5$. Then by Prop. 1.2.24), we may lift $A$ to a classical modular form $\widetilde{A}$ of weight $p-1$ over $\mathbf{Z}_p$. Although this lift is far from unique, the valuation of such a lift evaluated at a point $E/R$ does not depend on the lift providing that the valuation is $< 1$. It follows that:

$$\min\{1, v(\widetilde{A}(E/R, \omega_R)\} = \min\{1, v(A(E/S, \omega_S)\}$$

depends only on $E/S$. Hence, by abuse of notation, we may talk about the valuation of the Hasse invariant of $E$ and refer to this quantity. Explicitly, we may take $\widetilde{A} = E_{p-1}$. When $p = 2$ or 3, we may instead lift $A^4 \mod 8$ or $A^2 \mod 9$ to $E_4$ and $E_6$ respectively.

3.1. **Canonical subgroups for general $p$.** A key fact is the following generalization:

3.1.1. **Theorem** (Lubin–Katz)**.** *Let $R$ be a complete $\mathbf{Z}_p$-algebra. An elliptic curve $E/R$ has a canonical subgroup of order $p$ if and only if*

$$v(A) < \frac{p}{p+1},$$

*where $A(E_S, \omega_S)$ is the Hasse invariant of $E/S$ with $S = R/p$.*

We call elliptic curves $E/R$ satisfying the hypothesis of the theorem *not too supersingular*. How does one prove this for all $p$? They key point is to use formal groups, which (following [Sil86]) is a good way to understand elliptic curves over local fields.

Note that if $E$ is defined over $\mathbf{Z}_p$, then $E$ will never have a canonical subgroup unless $E$ is ordinary, because otherwise $v(A) \geq 1$. Suppose that $R$ is the ring of integers of a finite extension of $\mathbf{Q}_p$, and suppose that $E/R$ admits a canonical subgroup $C$. One might ask what the canonical subgroup looks like explicitly. It should define a finite flat group scheme over $\mathrm{Spec}(R)$ of order $p$. Yet such objects were classified by Oort–Tate ([TO70]), they take the form:

$$\mathrm{Spec}(R[x]/(x^p + \alpha \cdot x))$$

for some $\alpha, \beta \in R$ with $p = \alpha\beta$. Note that a change of variables allows one to change $\alpha$ by $\alpha\lambda^{p-1}$ and $\beta$ by $\lambda^{1-p}\beta$ for for any $\lambda \in R^\times$, and that this is the only isomorphism between these group schemes for distinct $\alpha$. $C$ is étale if and only if $\alpha \in R^\times$; Cartier duality has the effect of replacing $\alpha$ by $\beta \cdot w$ by some specific unit $w$ (see [TO70]).

3.1.2. **Theorem** (Coleman [Col05])**.** *The Canonical subgroup of $E$ is given by*

$$C = \mathrm{Spec}\left(\frac{R[x]}{\left(x^p + \dfrac{p}{A(E,\omega)} \cdot x\right)}\right).$$

Note that changing $\omega$ by a unit $\lambda$ does not change the isomorphism type of $C$.

3.2. **The curves $X^{\mathrm{rig}}[r]$.** (cf. [Buz03], and also [Con06].) Fix a modular curve $X$ of level prime to $p$, and assume that $X$ is a fine moduli space which is smooth over $\mathbf{Z}_p$. Let $k$ denote a finite extension of $\mathbf{F}_p$. The corresponding rigid analytic space $X^{\mathrm{rig}}$ admits a map

$$X^{\mathrm{rig}}(\mathbf{C}_p) \to X(\overline{k}).$$

The pre-image of any point $x$ is an open disc. The complement of the open discs corresponding to the supersingular points is the ordinary locus[5] $X^{\mathrm{rig}}[0]$. We would like to remove "smaller" discs. Let $x$ be a supersingular point, and let $E/k$ denote the corresponding elliptic curve. Since $X$ is smooth at $x$, the completion of $X$ at $x$ is isomorphic to $W[[t]]$, where $W = W(k)$ and $t$ is a local parameter. It is natural to define $X^{\mathrm{rig}}[r]$ by removing from $X^{\mathrm{rig}}$ the open balls $B$ of radius $|p^r|_p = p^{-r}$ in the parameter $t$. If $r = 0$, this recovers the ordinary locus $X^{\mathrm{rig}}[0]$.

3.2.1. **Lemma.** *This definition is independent of any choices provided that $r < 1$.*

*Proof.* Any different uniformizing parameter would be of the form $s = ap + ut$ where $a \in W$ and $u \in W[[t]]^\times$. Yet $v(s) = v(t)$ provided that either $v(s)$ or $v(t)$ is less than $v(p)$. □

3.2.2. *The canonical section.* The existence of the canonical subgroup produces a section of the natural map:

$$X_0^{\mathrm{rig}}(p) \to X^{\mathrm{rig}},$$

in a neighbourhood of $\infty$. Namely, we map $E$ to $(E, C)$ where $C$ is the canonical subgroup of $E$. For example, there is an isomorphism

$$X_0^{\mathrm{rig}}(p)[0] \to X^{\mathrm{rig}}[0],$$

[5]More precisely, there exists a rigid subspace $X^{\mathrm{rig}}[0] \subset X^{\mathrm{rig}}$ whose closed points are identified with the pre-image of the ordinary points over $\overline{k}$. However, from this point on, we shall elide the distinction between a rigid analytic space and its underlying set of closed points.

as long as we interpret $X_0^{\mathrm{rig}}(p)[0]$ to be the component of the ordinary locus containing $\infty$. Yet this section extends as far as the canonical subgroup can be defined, namely, to $X^{\mathrm{rig}}[r]$ for any

$$r < \frac{p}{p+1}.$$

To see this explicitly for $p = 2$, we need to recall some of the geometry of $X_0(2)$. It has genus zero, and thus it is given by the projective line for some modular function $f$. There are various choices of $f$ to make, but one classical one is the inverse of the Hauptmodul:

$$f = \frac{\Delta(2\tau)}{\Delta(\tau)} = q\prod_{n=1}^{\infty}(1+q^n)^{24} = q + 24q^2 + \dots$$

One has the classical modular equation:

$$\frac{f}{(1+2^8 f)^3} = j^{-1} = \left(\frac{1}{q} + 744 + \dots\right)^{-1} = q - 744q^2 + 356652q^3 \dots$$

The functions $f$ and $j^{-1}$ are both uniformizing parameters at the cusp $\infty$. Let us try to compute a section by solving the corresponding cubic equation:

$$(1+2^8 f)^3 - jf = 0$$

in a neighbourhood of $j^{-1} = 0$. The slopes (valuations of the coefficients) of this polynomial (as a polynomial in $f$) are

$$[24, 16, v(3\cdot 2^8 - j), 0].$$

In particular, as long as:

$$\|j^{-1}\|_2 < \|2^{-8}\| = 2^8,$$

there is a unique root $f$ of valuation $> -8$. How does this relate to our previous computation and Lemma 3.0.2? Note that

$$E_4 = 1 - 240\sum \sigma_3(n)q^n$$

is a lift of $A^4$. Moreover, we have

$$\frac{E_4^3}{\Delta} = j.$$

If we are close to the cusp of $X(1)$, then $\Delta$ is close to zero and there is a canonical subgroup corresponding to $\mu_p$ in $T(q)$ (The corrsponding elliptic curves have multiplicative reduction). Suppose instead that $E$ has good reduction at 2. Then, choosing $\omega$ so as to obtain a minimal model for $E$, we find that $\Delta(E,\omega)$ is a unit, and hence

$$v(E_4^3) = v(j).$$

In particular, the region $v(j) < 8$ corresponds (for curves of good reduction) to the region

$$v(A) = \frac{1}{4}v(E_4) = \frac{1}{12}v(E_4^3) < \frac{8}{12} = 2/3,$$

which is exactly the bound required to admit a congruence subgroup.

3.2.3. **Exercise.** *Show that, for the elliptic curve $E$ of exercise 3.0.3, one has $j = 2^6 + 2^9 + O(2^{11})$, and $f(E,C) = 2^{-6} + 2^{-4} + 2^4 + O(2^5)$.*

Let's also consider the case of $p = 37$. Because $p \equiv 1 \mod 12$, the $j$-invariants 0 and 1728 are ordinary, and hence there are exactly

$$\frac{p-1}{12} = 3$$

supersingular points, given explicitly by $j = 8$ and the roots $3 \pm \sqrt{15}$ of $\alpha^2 - 6\alpha - 6 = 0$. The ordinary locus $X^{\mathrm{rig}}[0]$ is simply the Riemann sphere minus three discs. Moreover, $X^{\mathrm{rig}}[r]$ will also be the Riemann sphere minus three slightly smaller discs. There is a map: $X^{\mathrm{rig}} \to [0,1]$ given by taking the minimum of 1 and the valuation of the Hasse invariant. The pre-image of $[0,r]$ may be identified with $X^{\mathrm{rig}}[r]$, by definition. Over any interval $[0,r]$ with $r < p/(p+1)$, there exists a section $s : X^{\mathrm{rig}}[r] \to X_0^{\mathrm{rig}}(p)$, which sends a $E$ to $(E, C)$, for the canonical subgroup $C$ of $E$. If $E$ is ordinary, then $C$ will be the kernel of the reduction map $E[p] \to E \to E(\overline{\mathbf{F}}_p)[p]$. It looks something like Figure 3.

3.3. **The reason everything works.** Suppose that

$$r < \frac{p}{p+1},$$

and consider the curve $X^{\mathrm{rig}}[r]$. If $(\ell, p) = 1$, then the Hecke operators $T_\ell$ extends to a correspondence on $X^{\mathrm{rig}}[r]$, since taking quotients by group schemes of order prime to $p$ does not effect the Hasse invariant. The key point, and literally everything hangs on this, is that, for a subgroup scheme $H$ of $E$ of order $p$ which is *not* the canonical subgroup, the valuation of $A(E/H, \phi^{\vee *}\omega)$ *decreases* as long as $0 < v(A) < p/(p+1)$. This is the key theorem:

3.3.1. **Theorem** (Katz–Lubin)**.** *Let $(E/R, \omega)$ be an elliptic curve and suppose that*

$$v(A(E,\omega)) < \frac{p}{p+1}.$$

*Suppose that $H \subset E$ is a subgroup scheme of order $p$ which is not the canonical subgroup. Let $\phi : E \to E/H$ be the natural projection, and $\phi^\vee : E/H \to E$ the dual isogeny.*

$$v(A(E/H, (\phi^\vee)^*\omega)) = \frac{v(A(E,\omega))}{p}.$$

The proof of this is not terribly hard, it requires knowing something about formal groups (which is mostly in [Sil86]), and is contained in [Kat73]. The identification of $X^{\mathrm{rig}}[r]$ with the component of $X_0^{\mathrm{rig}}(p)[r]$ containing $\infty$ allows us to define an operator $U_p$ on sections of $X^{\mathrm{rig}}[r]$; one simply takes the sum over all pairs $(E, H)$ where $H$ is *not* the canonical subgroup. As a consequence of the theorem above, we have the following:

3.3.2. **Theorem.** *Let $0 < r < 1/(p+1)$. Suppose that $f$ is a section of $H^0(X^{\mathrm{rig}}[r], \omega^k)$. Then $U_p f$ extends to a function on $H^0(X^{\mathrm{rig}}[pr], \omega^k)$. In particular, $U_p$ defines a map:*

$$U_p : H^0(X^{\mathrm{rig}}[r], \omega^k) \to H^0(X^{\mathrm{rig}}[pr], \omega^k).$$

*Proof.* Let $(E, \omega)$ denote an elliptic curve with $v(A(E,\omega)) < pr$. It suffices to show that we can extend $U_p f$ to $(E, \omega)$. By definition, to evaluate $f$ on $(E, \omega)$ involves evaluating $f$ on elliptic curves $E/P$ as $P$ runs over the $p$ subgroup schemes of $E[p]$ which are not the canonical subgroup. In particular, all those elliptic curves have Hasse invariant at most $r$, and thus $f$ is well defined. □

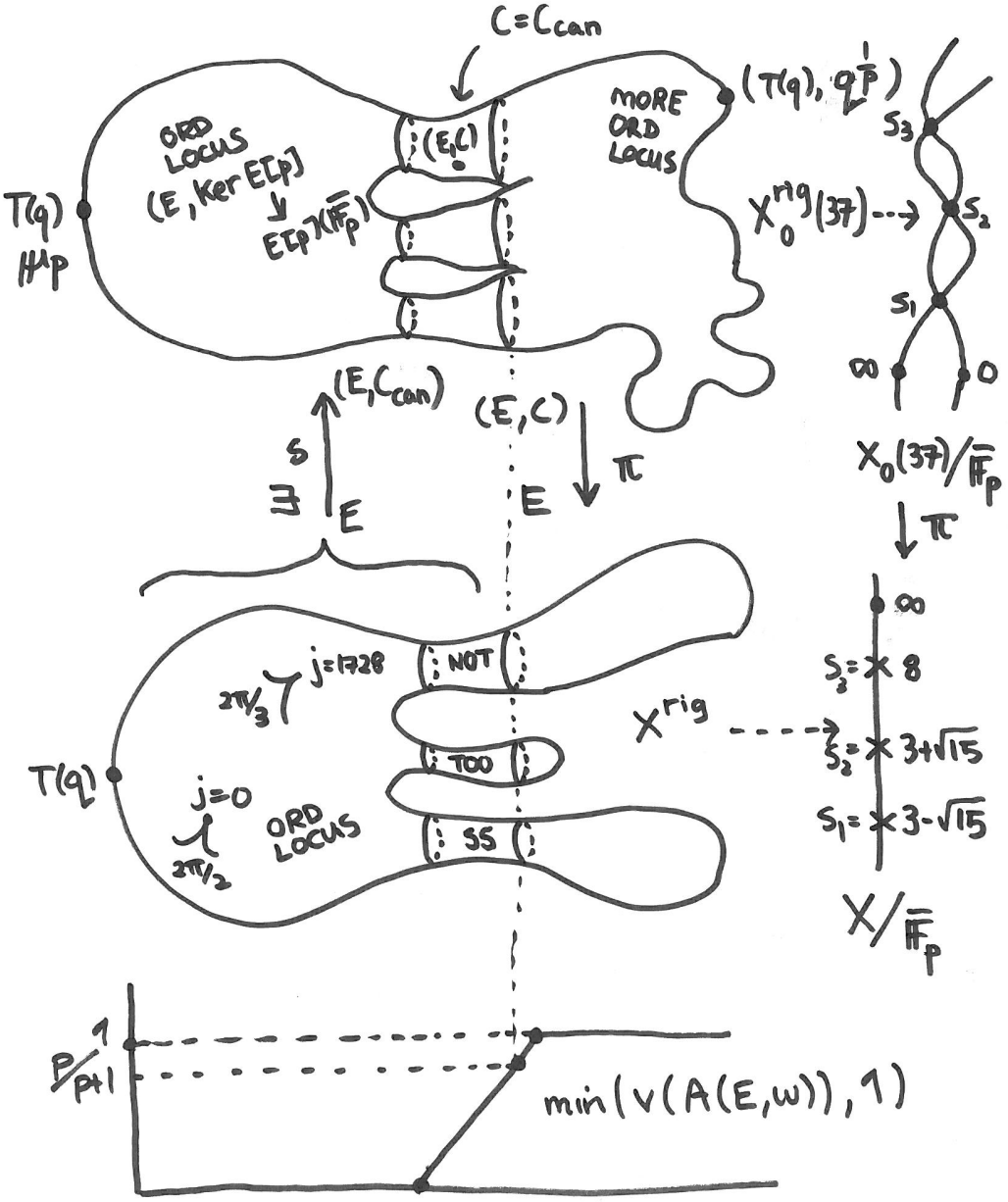


FIGURE 3. The map $X_0^{\mathrm{rig}}(37) \to X^{\mathrm{rig}}$ drawn as if $\mathbf{C}_{37}$ were archimedean

The correct way to think about this is that the operator $U_p$ *increases* the convergence of an overconvergent modular form. The next thing to consider is what type

of operators have this property. Imagine, for example, we let $C(r)$ denote the complex analytic functions on the closed ball $|z| \leq r$. Suppose we had an continuous operator

$$U : C(1) \to C(2).$$

Then we could compose $U$ with the restriction map $C(2) \to C(1)$. What is amazing about this last map is that it is compact.

3.4. **Overconvergent $p$-adic modular forms.**

3.4.1. **Definition.** *Let $0 < r < p/(p+1)$ be rational. The space of overconvergent modular forms of weight $k$, level $\Gamma$, and radius $r$, is defined to be:*

$$M_k^\dagger(\Gamma, r) = H^0(X^{\mathrm{rig}}[r], \omega^k).$$

3.4.2. **Remark.** *There is an inclusion:*

$$M_k^\dagger(\Gamma, r) \to M_k(\Gamma, \mathbf{C}_p, 0) \to \mathbf{C}_p[[q]].$$

*Hence overconvergent modular forms satisfy the $q$-expansion principle.*

3.4.3. **Example.** *Suppose that $N = 1$ and $p = 2$ and $k = 0$. Then*

$$M_0^\dagger(\Gamma, r) = \mathbf{C}_2\langle 2^r f\rangle$$

*is a ball of radius $2^{-r}$.*

3.4.4. **Lemma.** *$M_0^\dagger(\Gamma, r)$ is a Banach space with respect to the supremum norm on $X^{\mathrm{rig}}[r]$.*

Denote the norm by $\| \cdot \|_r$. The restriction maps;

$$\phi : M_0^\dagger(\Gamma, s) \to M_0^\dagger(\Gamma, r)$$

are continuous, since

$$\|g\|_s \leq \|\phi(g)\|_r.$$

The norm also makes sense when $r = 0$. In this case the forms are no longer *overconvergent* and thus we drop the $\dagger$.

3.4.5. **Exercise.** *Show that $\| \cdot \|_0$ co-incides with the $q$-expansion norm. Deduce that any sequence of overconvergent modular forms converging in $M_k^\dagger(\Gamma, r)$ are also converging in the $q$-expansion topology.*

It is not too difficult to construct sections of $\omega^k$ which don't vanish on $X^{\mathrm{rig}}[r]$, and hence $M_k^\dagger(\Gamma, r) \simeq M_0^\dagger(\Gamma, r)$ as Banach spaces for every integer $k$. Of course, these isomorphisms don't commute with Hecke operators.

3.5. **Compact operators and spectral expansions.** Let $U$ be a linear operator on a finite rank vector space $V$ (you can, if you wish, choose a basis for $V$ and think of $U$ as a matrix). Here we suppose that the coefficients lie in $\mathbf{R}$, or $\mathbf{C}$, or $\mathbf{Q}_p$, or $\mathbf{C}_p$, or any complete normed field $F$. The operator $U$ has $n$ generalized eigenvalues in some finite extension of $F$. For any $v \in V$, we may write

$$v = \sum \alpha_i v_i$$

for an eigenbasis $v_i$. Let us suppose that

$$|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n|.$$

That is, we assume that there is a unique greatest eigenvector. Then we observe that

$$\lim_{k\to\infty} \frac{U^k v}{\lambda_1^k} = \alpha_1 v_1.$$

If, on the other hand, we have

$$|\lambda_1| = |\lambda_2| = \ldots = |\lambda_m| > |\lambda_{m+1}| \geq \ldots \geq |\lambda_n|,$$

Then, if $\pi$ denotes the projection of $v$ onto the subspace of $V$ generated by $v_i$ for $i = 1$ to $m$, (so $\pi v = \sum \alpha_i v_i$ for $i \leq m$) we at least have:

$$\lim_{k\to\infty} \frac{U^k v}{\lambda^k} - \frac{U^k \pi v}{\lambda^k} \to 0,$$

where $\lambda$ is any of the eigenvalues $\lambda_i$ for $i \leq m$.

Now let us suppose that $V$ has infinite dimension. In order to make sense of continuity, we assume that $V$ has a norm, and is complete with respect to this norm; in particular, it is a Banach space. A random continuous linear operator need not have a spectrum, However, there exists a special class of operators, the *compact* operators, which admit a nice spectral theory (though not quite as nice as the finite dimensional case).

3.5.1. **Definition.** *A continuous bounded operator $U$ on a Banach space $B$ to itself is compact if the image of the unit ball is relatively compact.*

It turns out that compact operators are easier to understand in the ultrametric case because the norms are much easier to handle. Suppose that $B$ is a separable Banach space with an ultrametric norm (which will always be true in the cases we consider). Then being compact is equivalent to being a limit of operators of finite rank, which is equivalent to $U$ being a Nuclear operator (i.e. a compact operator such that the trace of $U$ and its powers are well defined). Note that in some sources (say in Coleman or in [Ser62]) these operators are called *completely continuous*, but we will not use that notation.

An operator $U$ as above admits (see [Dwo62], §2) a spectrum

$$|\lambda_1| \geq |\lambda_2| \geq \ldots \geq \ldots$$

where $|\lambda_i| \to 0$ as $i$ becomes arbitrarily large, and a sequence of generalized eigenvectors $v_1$, etc. such that any $v \in V$ admits an *asymptotic* expansion:

$$v \sim \sum \alpha_i v_i.$$

The asymptotic expression need not converge: consider, for example, the damped shift operator $U$ such that:

$$U x_{n-1} = \beta_n x_n$$

for a sequence $\beta_n$ such that $\lim \beta_n = 0$, and $\beta_n \neq 0$ for any $n$. Explicitly, we have

$$U(a_0, a_1, a_2, a_3 \ldots) = (0, \beta_1 a_0, \beta_2 a_1, \beta_3 a_2, \ldots),$$

and so on. Suppose that $Uv = \lambda v$. Let $a_{k-1}$ denote the first non-zero entry of $v$. Then the first non-zero entry of $Uv$ is $\beta_k a_k \neq 0$. Yet this contracts the equality $Uv = \lambda v$. Hence every $v \in V$ has a trivial asymptotic expansion. On the other hand, we have the following:

3.5.2. **Lemma** (Asymptotic Expansions)**.** *Suppose that $U$ acts compactly on a separable Banach space $B$ with an ultrametric norm. Then, for $v \in B$, there exists constants $\alpha_i$ and generalized eigenvectors $v_i$ of $U$ with non-zero eigenvalue $\sum \alpha_i v_i$ and a "spectral expansion" $v \sim \sum \alpha_i v_i$ with the following property. Let $\epsilon > 0$ be a fixed real number. Then, as $n$ goes to infinity,*

$$\left\| U^n v - \sum_{\geq \epsilon} \alpha_i U^n v_i \right\| = o(\epsilon^n),$$

*where the sum ranges over the finitely many generalized eigenvectors $v_i$ whose corresponding eigenvalue is $\geq \epsilon$.*

In particular, an asymptotic expansion allows one to understand $U^n v$ modulo any fixed power of $\epsilon$, with the necessary proviso that the implied error constants depend on $\epsilon$.

3.5.3. **Remark.** *Note that for a fixed eigenvalue $\lambda \neq 0$, the generalized eigenspace of $U$ is finite dimensional, but that not all generalized eigenfunctions may be* actual *eigenfunctions. This happens already in the finite dimensional case.*

3.5.4. **Exercise.** *Let $C(r)$ denote the complex analytic functions on $|z| \leq r$. Prove that the composition:*

$$C(1) \to C(2) \to C(1)$$

*defined by $Uf(z) = f(z/2)$ is compact. Determine all the eigenvectors of $U$, and prove that every element in $C(1)$ admits an absolutely convergent spectral expansion.*

The point of this exercise is that the map $U_p$ is exactly of the form, and hence:

3.5.5. **Theorem.** *Suppose that*

$$r < \frac{p}{p+1}.$$

*Then the map $U : H^0(X^{\mathrm{rig}}[r], \omega^k) \to H^0(X^{\mathrm{rig}}[r], \omega^k)$ is compact.*

The proof is that it is composed of a continuous map which extends convergence with the restriction map which is compact (this uses Theorem 3.3.2.) The big question then is, what type of compact operator is this?

3.6. **Classical Forms.** The following lemma is the analog of Lemma 2.1.6

3.6.1. **Lemma.** *There is an inclusion:*

$$M_k(\Gamma_0(p^n)) \subset M_k^\dagger(\Gamma, r)$$

*for any $k$ and small enough $r$.*

If $E/R$ is not too supersingular, then $E$ has a canonical subgroup $C$. As long as the Hasse invariant of $E/R$ is sufficiently large, we deduce that $E/C$ is also not too supersingular, and thus (by induction) as long as $r$ is sufficiently small, for suitable $E/R$ we may find a canonical subgroup $C$ of order $p^n$, from whence the lemma follows.

3.6.2. *Some important but not entirely relevant facts.* Suppose that $f$ is a classical eigenform of weight $k$ and level $\Gamma$. Suppose that $T_p$ has eigenvalue $a_p$. Consider the polynomial

$$x^2 - a_p x + p^{k-1},$$

which is the minimal polynomial of crystalline Frobenius (a fact which is both highly relevant and can be ignored completely). Associated to $f$ is a two dimensional space of old-forms of level $\Gamma_0(p)$, given explicitly by $f = f(q)$ and $U_p f = a_p f - p^{k-1} f(q^p)$. If $\alpha$ and $\beta$ are the roots of the characteristic polynomial (they are conjecturally distinct if $k > 1$) then $f(q) - \alpha f(q^p)$ and $f(q) - \beta f(q^p)$ have level $\Gamma_0(p)$ and are eigenvalues of $U_p$. They are overconvergent! Note that $v(\alpha), v(\beta) \leq p - 1$. There is (almost) a converse to this, namely, if $f$ is an overconvergent eigenform for $U_p$ with $U_p f = \lambda f$ and $v(\lambda) < k - 1$, then $f$ is classical. This is a theorem of Coleman [Col96]. When $v(\lambda) = k - 1$, it can (and does) go either way, although there are more refined conjectures predicting what should happen in this case.

3.7. **The characteristic power series.** Associated to the compact operator $U_p$ is the Fredholm power series $\det(1 - TU_p) \in \mathbf{Z}_p[[T]]$. Generalizing Hida's theorem, Coleman shows that as the weight varies, the coefficients of this series vary continuously in the weight. Moreover, they may be identified with elements in the Iwasawa algebra $\Lambda = \mathbf{Z}_p[[\mathbf{Z}_p^\times]]$. Using the fact that forms of small weight are classical, the usual trace formula allows one to give an exact formula for the coefficients of $\det(1 - TU_p)$ as finite sums involving class numbers. In particular, the coefficients are very *computable*, and thus, via Newton's Lemma, the valuations of the spectral eigenvalues $|\lambda_1| \geq |\lambda_2| \geq \dots$ are also very computable.

3.7.1. **Exercise.** *Show that any finite slope eigenvalue of $U_p$ lies in $M_k^\dagger(\Gamma, r)$ for any $r < \dfrac{p}{p+1}$.*

3.7.2. **Exercise.** *Prove that the trace of $U_2$ on $M_0^\dagger(1, r)$ is*

$$\frac{7 - \sqrt{-7}}{28} = 1 + 2^3 + 2^4 + 2^7 + 2^{10} + 2^{12} + 2^{13} + \dots$$

3.8. **The Spectral conjecture.** We have seen that, in general, the asymptotic expansion with respect to a compact operator need not be absolutely convergent. One may ask whether this sequence *does* converge in the special case of overconvergent modular forms with respect to the $U_p$-operator. One obstruction to convergence is as follows.

3.8.1. **Lemma.** *If the asymptotic expansion of an operator $U$ on a Banach space $B$ is convergent to the identity operator, then* $\ker(U) = 0$.

*Proof.* This is obvious. □

On the other hand, we have the following.

3.8.2. **Lemma.** *$V_p$ defines a map*

$$V_p : M_k^\dagger(\Gamma, r) \to M_k^\dagger(\Gamma, r/p).$$

The proof is virtually the same as the proof that $V_p$ preserves $p$-adic modular forms. More precisely, $V_p f$ evaluated on $E$ depends only on $f$ evaluated at $E/C$,

where $C$ is the canonical subgroup. Yet this *increases* the valuation of the Hasse invariant. On $q$-expansions, we have

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

In particular, the composition $U_p V_p$ is the identity. (This follows from the $q$-expansion principle.) Let $W_p$ be the operator $1 - V_p U_p$. Then $U_p W_p = U_p - U_p V_p U_p = U_p - U_p = 0$. In particular, if $f$ lies in $\ker(U_p)$ then $W_p f = f$, and moreover, the image of $W_p$ lies in the kernel of $U_p$. On $q$-expansions, we have

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

3.8.3. **Lemma.** *Suppose that* $r < \dfrac{1}{p+1}$*, then* $W_p$ *defines a map:*

$$M_k^\dagger(\Gamma, r) \to M_k^\dagger(\Gamma, r)$$

*which is a projection onto* $\ker(U_p)$*.*

The reason the bound on $r$ is needed is that for larger $r$, it is not necessarily the case that $U_p$ increases the radius of convergence by a factor of $p$, and thus the composite $V_p U_p$ may decrease the radius of convergence. We immediately deduce from this the following:

3.8.4. **Lemma.** *Suppose that*

$$r < \frac{1}{p+1}.$$

*Then the kernel of* $U_p$ *on* $M_k^\dagger(\Gamma, r)$ *is infinite dimensional. In particular, the spectral expansion of* $U_p$ *for such* $r$ *is not in general convergent.*

On the other hand, there seems to be a transition that takes place at $r = 1/(p+1)$, as indicated by the following lemma.

3.8.5. **Lemma.** *If* $r > \dfrac{1}{p+1}$*, then the kernel of* $U_p$ *on* $M_k^\dagger(\Gamma, r)$ *is trivial.*

*Proof.* This is [BC06] Lemma 6.13 (and Remark 6.14). □

The kernel of a compact operator is not the only obstruction to convergence. Recall that the damped shift operator $U x_{n-1} = \beta_n x_n$ considered above (where $\lim \beta_n = 0$ is a sequence of non-zero elements) has trivial spectral expansions even though $U$ itself has no kernel. There still, however, appears to be reason to believe the following.

3.8.6. **Conjecture.** *Suppose that* $r \in (1/(p+1), p/(p+1))$*. Then any* $F \in M_k^\dagger(\Gamma, r)$ *has a convergent spectral expansion which converges to* $F$*.*

Explicitly, we may write any $F \in M_k^\dagger(\Gamma, r)$ as

$$F = \sum \pi_i(F) v_i,$$

where the $v_i$ are a fixed choice of (generalized) eigenvectors with eigenvalues $\lambda_i$. Note that by "convergence" above we mean convergence in the Banach space norm on $M_k^\dagger(\Gamma, r)$ (that is, the supremum norm). This is a much more restrictive condition than convergence in the $q$-expansion topology (which is the supremum norm on the ordinary locus $X^{\mathrm{rig}}[0]$.)

3.8.7. **Remark.** *Since the norm on $X^{\mathrm{rig}}[r]$ is non-Archimedean, given any spectral expansion of $F$ as above one has:*

$$\|F\|_r \leq \sup |\pi_i(F)| \cdot \|v_i\|_r.$$

It is natural to supplement the spectral conjecture with the guess that $\|F\|_{1/2} = \sup |\pi_i(F)| \cdot \|v\|_{1/2}$, which would be a consequence of knowing that the eigenvectors $v_i$ are sufficiently disjoint.

This conjecture also has immediate consequences for a form $F \in M_k^\dagger(\Gamma, r)$ for *any* $r$.

3.8.8. **Lemma.** *Let $F \in M_k^\dagger(\Gamma, r)$, and suppose that $F \sim \sum \alpha_i v_i$ is the asymptotic expansion of $F$. Assume Conjecture 3.8.6. Then*

$$U_p^n F = U_p^n \sum \alpha_i v_i.$$

*for sufficiently large $n$. One may take any $n$ such that $v(r^n) > 1/(p+1)$.*

3.8.9. **Remark** (Remark on semisimplicity of $U_p$)**.** One might wonder if $F$ can actually be decomposed into *eigenfunctions*. It turns out that this is a subtle question even for classical forms. The action of $U_p$ on the two dimensional space of old forms for an eigenform $f$ of level prime to $p$ is given by (with respect to one basis):

$$\begin{pmatrix} a_p & p^{k-1} \\ -1 & 0 \end{pmatrix}.$$

This is semi-simple only if $a_p^2 \neq p^{k-1}$. It is still unknown whether this can happen, although it follows from the Tate conjecture [CE98]. One certainly expects — even if $U_p$ fails to be semi-simple — that the corresponding generalized eigenforms all decompose into actual eigenforms for the Hecke operators $T_\ell$ with $\ell$ prime to $p$ and the level.

3.9. **The invariant pairing.** How does one prove that a spectral expansion of a compact operator $U$ on a Banach space $B$ exists and is convergent? A natural way is to show that the operator $U$ actually preserves extra structure, namely, that $B$ has the structure of a Hilbert space $H = (B, \langle , \rangle)$ such that $U$ is self-adjoint. One of the problems with trying to apply this to our case is that there is no notion of Hilbert space for non-Archimedean fields. The point is that quadratic forms in sufficiently many variables over $\mathbf{Q}_p$ are never anisotropic. (That is, quadratic forms have zeros.) It follows that it's very hard to define a "non-degenerate" quadratic form, since one will invariably end up with vectors $v$ such that $\langle v, v \rangle = 0$. On the other hand, it turns out that the operator $U_p$ on $M_0^\dagger(\Gamma, r)$ *does* preserve a natural pairing, as long as $r$ is sufficiently big.

Let $(E/R, \omega_R)$ be an elliptic curve with

$$v(A(E, \omega)) = s < \frac{p}{p+1}.$$

We know that $E$ admits a canonical subgroup $C$, and one has a corresponding point $(E, C)$ on $X_0^{\mathrm{rig}}(p)$, the image of $E$ under the section $X^{\mathrm{rig}}[r] \to X_0^{\mathrm{rig}}(p)$ for $r > s$. The Fricke involution $w_p$ acts on $X_0^{\mathrm{rig}}(p)$ by sending $(E, C)$ to $(E/C, E[p]/C)$. If $E$ is the Tate curve $T(q)$, for example, then $(E, C) = (T(q), \mu_p)$ and $(E/C, E[p]/\mu_p) =$

$(T(q^p), \{q\})$. In particular, the corresponding subgroup is no longer the canonical subgroup, and $E/C$ is not in the image of $X^{\mathrm{rig}}[r]$. If, however,

$$\frac{1}{p+1} < s < \frac{p}{p+1},$$

then $E[p]/C$ *has* a canonical subgroup which may be identified with $E/C$ ([Buz03]). Moreover (*ibid.*) one has an equality

$$v(A(E/C, \phi^*\omega)) = 1 - s.$$

It follows that, if $v \in M_k^\dagger(\Gamma, r)$ then $w_p^* v$ is a function defined on pairs $(E, \omega)$ such that

$$1 > v(A(E, \omega)) \geq 1 - r.$$

Suppose that $r \geq 1/2$. Then for any pair of functions $u$ and $v$ in $M_0^\dagger(\Gamma, r)$, both $u$ and $w_p^* v$ are then both defined on the annulus $|t| = |p^{1/2}|$, where $t$ is a local parameter at the supersingular point. In particular, as long as $r \geq 1/2$, one may define a pairing on $M_0^\dagger(\Gamma, r)$ as follows:

$$\langle u, v \rangle = \int w^* v du := \mathrm{Res}_{z=\infty} w^* v du.$$

3.9.1. **Lemma** (Loeffler [Loe07])**.** *This pairing is $U_p$ and Hecke equivariant.*

If $N = 1$ and $p = 2$ and $r = 1/2$, then $M_0^\dagger(1, 1/2) = \mathbf{C}_2 \otimes \mathbf{Z}_2[[g]]$, where $g = 2^6 f$. Note that $w^* g = g^{-1}$, and so

$$\langle g^m, g^n \rangle = \int g^{-m} \cdot n g^n \frac{dg}{g} = \begin{cases} m, & m = n \\ 0, & m \neq n. \end{cases}$$

3.9.2. *Symmetric operators.* A symmetric matrix over $\mathbf{C}$ is not necessarily diagonalizable. One might ask if being symmetric allows one to deduce *anything*. Let $B$ be a Banach space over $\mathbf{C}_p$ with $|B| = p^{\mathbf{Q}} \cup \{0\}$. Suppose that $B$ admits a continuous bilinear pairing

$$\langle \cdot, \cdot \rangle : B \times B \to \mathbf{C}_p.$$

Suppose, furthermore, that $B$ admits a topological basis $\{x_i\}$ such that

$$\langle x_i, x_j \rangle = \delta_{ij}.$$

3.9.3. **Question.** *Let $U$ be a compact operator on $B$ that is equivariant with respect to the pairing, that is,*

$$\langle Ux, y \rangle = \langle x, Uy \rangle.$$

*Suppose that $B$ contains a non-zero vector $v$ such that the action of $U$ on the closure of the vectors $U^n v$ for all $n$ is topologically nilpotent. Then is it the case that* $\ker(U) \neq 0$*?*

3.10. **A special case of the spectral conjecture.** One piece of evidence for this spectral conjecture is the following.

3.10.1. **Theorem** (Loeffler [Loe07])**.** *The spectral conjecture is true if $N = 1$ and $p = 2$ for $r \in (5/12, 7/12)$. Moreover, one also has $\|F\|_r = \sup \|\alpha_i \phi_i\|_r$ in that range.*

In light of the main theorem of [BC05], one has the following.

3.10.2. **Theorem.** *Let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots$ denote the eigenvalues which occur for $N = 1$ and $p = 2$ of valuation $3, 7, 13, \dots$. Let $v_i$ denote the corresponding eigenform, normalized so that the leading coefficient is 1. If $F \in M_0^\dagger(1, 1/2)$, then there is an equality*

$$F = \sum \alpha_i \phi_i, \text{ where } \alpha_i := \frac{\langle F, \phi_i \rangle}{\langle \phi_i, \phi_i \rangle}.$$

*Moreover, $\|\alpha_i \phi_i\|_{1/2} \leq \|F\|_{1/2}$ for all $i$.*

3.10.3. **Lemma.** *If $\phi_i$ is a normalized overconvergent eigenform, then*

$$\|\phi_i\|_r \geq 1.$$

*Proof.* We noted previously that $\|\phi_i\|_r \geq \|\phi_i\|_0$. The latter is given by the $q$-expansion norm, and thus (since $\phi_i$ is normalized) it follows that $\|\phi_i\|_0 = 1$. □

By Lemma 3.10.3, we deduce that

$$|\alpha_i| \leq \frac{\|F\|_{1/2}}{\|\phi_i\|_{1/2}} \leq \|F\|_{1/2}.$$

In practice, one expects $\|\phi_i\|_{1/2}$ to increase relatively quickly. However, this estimate at least allows for an explicit computation of $\alpha_i$. We return to the numerology of eigenforms in section §5.

3.11. **Some heuristics.** Let us now reformulate the spectral conjecture in a slightly different way in weight 0. First, suppose we are working with classical cusp forms in $S_k(\Gamma, \mathbf{C})$. Then, for a cusp form $F$, one has an identity:

$$F = \sum \frac{\langle \phi_i, F \rangle}{\langle \phi_i, \phi_i \rangle} \phi_i,$$

where the right hand side is a *finite sum* over cuspidal eigenforms $\phi_i$, and $\langle *, * \rangle$ is the Petersson inner product, given by

$$\langle \phi, \psi \rangle = \int_\Omega \phi \overline{\psi} \cdot y^k \frac{dxdy}{y^2},$$

which satisfies

$$\langle \phi, \phi \rangle = L(1, \mathrm{ad}^0 \phi)$$

for eigenforms $\phi$. On the other hand, we expect that for $F \in S_0^\dagger(\Gamma, r)$ (and $r$ sufficiently large), one has an identity:

$$F = \sum \frac{\langle \phi_i, F \rangle}{\langle \phi_i, \phi_i \rangle} \phi_i,$$

where the right hand side is now an *infinite sum* over finite slope eigenforms $\phi_i$, and $\langle *, * \rangle$ is the invariant pairing described above.

3.11.1. **Exercise** ($\star$)**.** *Show that, suitably normalized, the invariant pairing $\langle \phi, \phi \rangle$ for a finite slope eigenform $\phi$ coincides with the $p$-adic $L$-function $L_p(1, \mathrm{ad}^0 \phi_\kappa)$ at $\kappa = 0$, where $\phi_\kappa$ denotes the Coleman family of eigenforms of weight $\kappa$ passing through $\phi$.*

Here are some thoughts on this exercise. Note that the $p$-adic adjoint $L$-function is related to the ramification of the Coleman family $\phi_\kappa$ over weight space. In particular, $L_p(1, \mathrm{ad}^0\phi_\kappa)$ should have zeros exactly at the ramification points (results of this flavour were proved by Kim in his thesis [Kim06]). On the other hand, assuming the existence of spectral expansions one expects that $\langle \phi, \phi \rangle = 0$ where $U\phi = \lambda\phi$ if and only if there exists a generalized eigenform $\psi$ such that $(U-\lambda)\psi = \phi$ (see the calculation of §5.0.5 for one direction, and use the $q$-expansion principle and the fact that $\langle *, \phi \rangle$ is non-vanishing for the other direction). Yet the non-semi-simplicity of $U$ is equivalent to the eigencurve being ramified at this point.

## 4. Examples

In this section, we give some explicit examples in order to illustrate the general theory. Write

$$j = \frac{1}{q} + 744 + 196884q + \ldots = \sum c(n)q^n.$$

We first show how to understand congruences for $c(n)$ modulo powers of two using a classical method, and we shall return later and use a modern approach, which gives more information.

4.1. **An example: $N = 1$ and $p = 2$; the Watson approach.** Recall that $X_0(2)$ is uniformized by the function:

$$f = q\prod_{n=1}^{\infty}(1+q^n)^{24} = q + 24q^2 + \ldots$$

and that there is an identity

$$\frac{(1+2^8 f)^3}{f} = j.$$

We first apply $U_2$ to $j$, and we find that:

$$U_2 j = 744 + \sum c(2n)q^n.$$

Formally, $U_2$ takes functions on $X_0(1)$ to $X_0(2)$. Thus $U_2 j$ is a meromorphic function on $X_0(2)$. Moreover, since $U_2 j(E)$ is a sum of $j(E/C)$ for various $C$, the function $U_2 j$ will be holomorphic on $X_0(2)$ away from the cusps. Since $U_2 j$ is holomorphic at $\infty$, it can only have poles at the other cusp of $X_0(2)$, namely at $f = \infty$, and hence $U_2 j$ is a polynomial in $f$. Indeed:

$$\begin{aligned} U_2 j - 744 &= 140737488355328 f^4 + 3298534883328 f^3 + 19730006016 f^2 + 21493760 f \\ &= 2^5\left(262144 g^4 + 393216 g^3 + 150528 g^2 + 10495 g\right), \end{aligned}$$

where $g = 2^6 \cdot f$. On the other hand, if $h$ is a meromorphic function on $X_0(2)$ then so is $U_2 h$, and if $h$ only has a pole at 0 then so does $U_2 h$; that is, $U_2$ takes polynomials in $f$ to polynomials in $f$. We see:

$$U_2 f = 24f + 2048f^2,$$

$$U_2 f^2 = f + 1152f^2 + 196608f^3 + 8388608f^4,$$

and so on. More generally,

$$U_2 f^n = \frac{1}{2}\left(f\left(\frac{\tau}{2}\right) + f\left(\frac{\tau+1}{2}\right)\right).$$

Hence $U_2 f^n$ satisfies a recurrence relation $x_n - a_1 x_{n-1} + a_2 x_{n-2} = 0$, where

$$X^2 - a_1 X + a_2 = \left(X - f\left(\frac{\tau}{2}\right)\right)\left(X - f\left(\frac{\tau+1}{2}\right)\right) = X^2 - (48f + 4096f^2)X - f.$$

The classical idea is now to *explicitly* compute the "matrix" of $U$ on some nice basis. If, for example, one shows that this matrix is divisible by 8 (in this case), then iterating $U$ will establish the necessarily congruences.

4.1.1. **Lemma.** *Let $h = 8f$, and consider the ring $R = \mathbf{Z}_2[[h]]$ of power series in $h$ with integral coefficients. Then the operator $\mathcal{F} := U_2/8$ acts continuously on $h \cdot R$.*

*Proof.* Continuity is equivalent to asking that $\mathcal{F}(h^n) \in R$, and the degree of $U_2 h^n$ goes to infinity with $n$. Both claims follow by induction. From the computations above, we see that:

$$\mathcal{F}(h) = 3h + 32h^2, \qquad \mathcal{F}(h^2) = h + 144h^2 + 3072h^3 + 16384h^4,$$

and then $\mathcal{F}(h^n) = 16(3h + 32h^2)\mathcal{F}(h^{n-1}) + 8h\mathcal{F}(h^{n-2})$. □

Since $U_2 j - 744 \in 2^8 h \cdot R$, it follows that

$$\sum_{n=1}^{\infty} c(2^m n) q^n = U_2^m j - 744 = (8\mathcal{F})^{m-1}(U_2 j - 744)$$
$$\subset (8\mathcal{F})^{m-1}(2^8 h \cdot R) \subset 2^{3m+5}\mathcal{F}(h \cdot R) \subset 2^{3m+5} h \cdot R \subset 2^{3m+8}\mathbf{Z}_2[[q]].$$

This proves Lehmer's congruence in the introduction.

4.2. **An example: $N = 1$ and $p = 2$; the Coleman approach.** The function $j$ defines a meromorphic function on $X^{\mathrm{rig}}[r]$ with a pole only at $\infty$, and hence $U_p j$ extends to an element of $M_0^\dagger(\Gamma, r)$ for any

$$r < \frac{p}{p+1}.$$

The operator $U_p$ on this space is *compact*. Now let $p = 2$. We may manually compute the first few slopes of the spectrum of $U_2$ to be 0, 3, 7, and 13. Of course, 1 is an eigenvalue for $U_2$ with slope zero. In particular, for *any* overconvergent form $g$ in $M_0^\dagger(\Gamma, r)$ with no constant term we have, from the asymptotic expansion Lemma 3.5.2), that

$$g \sim \alpha_1 \phi_1 + \alpha_2 \phi_2 + \alpha_3 \phi_3 + \dots$$

and thus:

$$U_2^m(U_2 j - 744) = \alpha_1 \lambda_1^m \phi_1 + \alpha_2 \lambda_2^m \phi_2 + o(2^{13m}).$$

How may one compare these arguments? The Watson style argument essentially proves *by hand* that $U_2$ is compact, and indeed that the norm of $\mathcal{F} = U_2/8$ on the cuspidal overconvergent forms is 1. This justifies the claim in the introduction — Coleman gives you the compactness of $U$ by geometry, whereas Watson gives it to you by *explicit computation*, but by a computation which needs to be redone every single time to get the best bounds. Moreover, such computations become essentially infeasible as soon as $X_0(p)$ has genus $> 0$. On the other hand, the fact that $\mathcal{F}$ has operator bound 1 is is *stronger* than the fact that the first eigenvalue has slope 3, even if it doesn't say anything about the higher order eigenvalues. How may we reconcile these two approaches?

Let us see what can be extracted from the spectral conjecture, which is a theorem in this case. We may write

$$U_2 j - 744 = \sum \alpha_i \phi_i, \text{ where } \alpha_i := \frac{\langle U_2 j, \phi_i \rangle}{\langle \phi_i, \phi_i \rangle}.$$

Moreover, $\|\alpha_i\|_{1/2} \leq \|U_2 j - 744\|_{1/2}$ for all $i$. Given the formula for $U_2 j$ above, and the fact that $\|g\|_{1/2} = 2^6 \|f\|_{1/2} = 1$, we deduce that there is an identity $\|U_2 j\|_{1/2} = |2^5| = 2^{-5}$, and thus $v(\alpha_i) \geq 5$. It follows that

$$U_2^m (U_2 j) = \alpha_1 \lambda_1^m \phi_1 + \alpha_2 \lambda_2^m \phi_2 \mod 2^{13m+5},$$

that is, we have made the constant above *effective*. From this we may compute easily enough that $\alpha_1 = 2^{11} + 2^{12} + \ldots$ and $\alpha_2 = 2^{16} + 2^{17} + \ldots$ from which we deduce the congruence of Lehmer. Yet we see that we get something much stronger, namely, that not only does $v(c(2^m)) = 3m + 8$, but

$$\sum_{n=1}^{\infty} \frac{c(2^m n)}{c(2^m)} q^n \mod 2^{4m+1}$$

is a Hecke eigenform. More generally, we have the following:

4.2.1. **Lemma.** *Let $F \in M_0^\dagger(1, 1/2)$, and suppose that $F$ is normalized so that $\|F\|_{1/2} = 1$. Then the spectral expansion takes the form:*

$$2^5 \cdot \sum \alpha_i \phi_i$$

*where $\alpha_i$ is divisible by $2^i$.*

*Proof.* This follows from the estimates on $\|\phi_i\|_{1/2}$ we shall prove in Lemma 5.1.9 □

4.3. **An example: the coefficients of $c(n)$ modulo powers of $p$.** Let's now consider a more general example, which seems harder to prove by any direct computation.

4.3.1. **Example.** *Let $j = \frac{1}{q} + 744 + 196884q + \ldots = \sum c(n) q^n$ be the modular $j$ invariant. Let $\widetilde{j} := e_p j$ denote the projection of $j$ to the ordinary subspace. Then there exists a constant $c$ depending only on $p$ such that*

$$U^n j \equiv U^n \widetilde{j} \mod p^{n-c}.$$

*Proof.* We first prove that there does not exist an overconvergent eigenform of weight zero and slope $\alpha$ with $0 < \alpha < 1$. Assume otherwise. Then, by theory of Coleman, there exists a *classical* form with the same slope and (possibly very large) weight $0 \mod p-1$. By Theorem 1.6 of [BG09], it follows that if $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ is the corresponding mod-$p$ Galois representation attached to this form, and $D_p = \mathrm{Gal}(\overline{\mathbf{Q}}_p / \mathbf{Q}_p) \subset G_{\mathbf{Q}}$ is the decomposition group at $p$, then

$$\omega^2 \otimes \overline{\rho} | D_p = \omega^2 \otimes \mathrm{Ind}_K^{\mathbf{Q}_p} \omega_2^{p-2} = \omega_2^{2p+2} \otimes \mathrm{Ind}_K^{\mathbf{Q}_p} \omega_2^{p-2} = \mathrm{Ind}_K^{\mathbf{Q}_p} \omega_2^{3p} = \mathrm{Ind}_K^{\mathbf{Q}_p} \omega_2^3.$$

In particular, it follows that $\overline{\rho} | D_p$ and thus $\overline{\rho}$ is irreducible (as long as $p \neq 2$, which we already considered), and hence, by the weight part of Serre's conjecture, the twist $\omega^2 \otimes \overline{\rho}$ gives rise to a non-trivial class in

$$S_4(\mathrm{SL}_2(\mathbf{Z}), \mathbf{F}_p) = 0,$$

a contradiction. In particular, the eigenfunction of weight zero with largest slope which is not ordinary has slope at least 1. The result then follows immediately from the asymptotic expansion. □

4.3.2. **Remark.** *Note that exactly the same argument — and conclusion — applies to* any *overconvergent p-adic modular form of weight* 0 *and level* 1*.*

4.3.3. **Exercise.** *Let $g$ be any p-adic overconvergent modular function which is congruent to* $1 \mod p$*, for example,* $g = 1 + p\dfrac{\Delta(p\tau)}{\Delta(\tau)}$*. Prove that*

$$g^s := \exp(s\log(g)) = 1 + s(g-1) + \binom{s}{2}(g-1)^2 + \binom{s}{3}(g-1)^3 + \dots$$

*is also overconvergent for sufficiently small $s \in \mathbf{C}_p$. Compute what sufficiently small means explicitly in this case.*

A natural question that presents itself is as follows: Can one effectively compute the constant $c$? Suppose one assumed the existence of a *convergent* spectral expansion, together with the estimate $\|F\|_r = \sup |\alpha_i| \cdot \|\phi_i\|_r$ for $r = 1/2$. As with $p = 2$, we would then have:

$$U_p j = U_p \widetilde{j} + \alpha_1\phi_1 + \alpha_2\phi_2 + \dots$$

and it would suffice to obtain effective and uniform bounds for $\alpha_i$. Yet there are obvious bounds $\|\phi_i\|_r \geq 1$ and $\|U_p j\|_r \leq p$ for all $r$, and thus $|\alpha_i| \leq p$. As we shall see later, it is most likely the case that the norms $\|\phi_i\|_r$ grow extremely rapidly (exponentially in $i$) and thus the $\alpha_i$ decrease to zero in a concomitant fashion.

4.3.4. **Exercise** ($\star$)**.** *What is the optimal upper bound for $\|U_p j\|_r$ for general $p$? What about the optimal upper bound for the operator norm $\|U_p\|_r$?*

4.4. **An example: convergence slower than $O(p^n)$.** There do exist forms of slope strictly between 0 and 1, which may effect the rate of convergence. To give an easy example, let

$$f = \sum a(n)q^n = q\prod_{n=1}(1-q^n)^2(1-q^{11n})^2 \in S_2(\Gamma_0(11), \mathbf{Z}),$$

which corresponds to the modular curve $X_0(11)$. This elliptic curve has supersingular reduction at 2, and the minimal polynomial of crystalline Frobenius is $x^2 + 2x - 2$. It follows that the corresponding old forms $f_\alpha$, $f_\beta$ with $\alpha, \beta = -1 \pm \sqrt{3}$ of level $\Gamma_0(22)$ each have slope $1/2$. In particular, it is not too hard to show that:

$$\sum a(2^m n)q^n \equiv 0 \mod 2^{\lceil \frac{m}{2} \rceil},$$

but that there is no such congruence modulo any higher power of 2. Of course, the same thing happens (with the same form) for any of the infinitely many primes $p$ such that $X_0(11)$ is supersingular.

4.4.1. **Exercise.** *Show that if*

$$\frac{E_4\Delta}{E_{58}} = \frac{\left(1 + 240\sum \sigma_3(n)q^n\right) q\prod_{n=1}^{\infty}(1-q^n)^{24}}{\left(1 - \dfrac{1416}{291322804651310489179471641358744 9}\sum \sigma_{58}(n)q^n\right)} =: \sum d(n)q^n,$$

*then $\sum d(59^m n)q^n$ converges to zero no faster than $O(59^{m/2})$.*

4.5. **Forms of half integral weight.** (cf. [Ram06, Ram08]). One may ask whether there exists a corresponding theory of $p$-adic and overconvergent modular forms of *half*-integral weight. The answer is yes. First recall how modular forms of half integral weight are defined — one starts with a *particular* modular form $\theta = \sum q^{n^2}$ and uses it to define (analytically) a square root of the sheaf $\omega$. On the other hand, the form $\theta$ certainly lies in $\mathbf{Z}[[q]]$, and so with a little care one can carry out these constructions a little more arithmetically. A key point in Coleman's work is that, as far as the analysis goes, one can pass between any integral weight $k \equiv 0 \mod p-1$ and weight 0. In particular, suppose that $k$ is positive. Then there is an isomorphism of Banach spaces:

$$\psi : M_k^\dagger(\Gamma, r) \to M_0^\dagger(\Gamma, r)$$

defined by divisiion by $V_pE_k$. (Since $E_k \equiv 1 \mod p$, it doesn't vanish on the ordinary locus, and hence for formal reasons both $E_k$ and $V_pE_k$ are invertible for some $r > 0$ — one can be more explicit.) Although this map is not $U_p$-equivariant, one may define a *twisted* operator by the formula:

$$\widetilde{U} = \frac{E_k}{V_pE_k} \cdot U.$$

Then, for $F$ of weight $k$, one has

$$\psi(U_pF) = \frac{U_pF}{V_pE_k} = \frac{E_k}{V_pE_k} \cdot \frac{U_pF}{E_k} = \frac{E_k}{V_pE_k} \cdot U\frac{F}{V_pE_k} = \widetilde{U}\psi(F).$$

The key observation, however, is that one may now extend this to any weight $\kappa$ in $\mathrm{Hom}(\Lambda, \mathbf{C}_p)$, by replacing $E_k$ by the $p$-adic Eisenstein series:

$$E_\kappa = \frac{\zeta_p(\kappa)}{2} + \sum_{n=1}^{\infty} \left( \sum_{d|n} \frac{\kappa(d)}{d} \right) q^n.$$

When $\kappa$ is the weight $x \mapsto x^k$, one recovers the classical Eisenstein series $E_k^*$ of weight $\Gamma_0(p)$. For forms of half-integral weight, one may also transfer the actions of Hecke operators to any other weight in exactly this way. Note that for half-integral weights, the appropriate operator at $p$ is $U_{p^2}$.

4.5.1. **Theorem** (Ramsey)**.** *Let $k$ be a half-integer. There exists a space of overconvergent forms $M_k^\dagger(\Gamma, r)$ of weight $k$, and $U_{p^2}$ acts compactly on this space.*

4.6. **An example: congruences for $p(n)$ modulo powers of $p$.**

4.6.1. **Example.** *Let*

$$\eta^{-1} = \frac{1}{q^{1/24}} + q^{23/24} + 2q^{47/24} + \ldots = \sum p\left(\frac{n+1}{24}\right) q^n$$

*be the inverse of Dedekind's eta function. Let $\widetilde{\eta^{-1}}$ denote the projection of $\eta^{-1}$ to the ordinary subspace. Then there exists a constant $c$ depending only on $p$ such that*

$$U^n\eta^{-1} \equiv U^n\widetilde{\eta^{-1}} \mod p^{n-c}.$$

Note that this has no content if $p = 2$ or $p = 3$, so assume that $p > 3$.

The form $\eta^{-1}$ is meromorphic of weight $-1/2$. The form $\eta^{-1}(24\tau)$ has level $\Gamma_0(576)$ and character $\chi$ — but it will be relevant to note that $\eta$ has extra symmetries — suggested, for example, by the fact that $\eta^{24}$ has level one. In particular,

one may define modular forms of half integral weight in a different way — by using $\eta$ instead of $\theta$.

4.6.2. **Definition.** *The modular forms $M_k(1, \mathbf{C})$ of half-integral weight $k$ and $\eta$-level one are the holomorphic forms on $\mathbf{H}$ which are bounded at the cusps and such that:*

$$\frac{f(\gamma\tau)}{f(\tau)} = \frac{\eta(\gamma\tau)^k}{\eta(\tau)^k}.$$

We can't quite use this to give a splitting of $\omega$ at level one, for stacky reasons. However, by allowing various different auxiliary levels this space admits a good integral structure[6]. Note that $\omega^{\otimes 12}$ *does* exist on $X(1)$, and that there are natural maps:

$$M_{12}(1, \mathbf{C}) \otimes M_k(1, \mathbf{C}) \to M_{12+k}(1, \mathbf{C}).$$

where $k$ is half-integral. We have

$$U_{p^2}\eta^{-1} \in S^\dagger_{-1/2}(1, r).$$

Alternatively, in the usual normalization, we certainly have

$$U_{p^2}\eta^{-1}(24\tau) \in S^\dagger_{-1/2}(\Gamma_0(576), r).$$

Let $e$ denote Hida's idempotent operator, and let $F = U_{p^2}\eta^{-1}(24\tau)$. Then, formally, there is an equality

$$F = e_p F + H + H_{\geq 1},$$

where $H$ is a finite sum of generalized eigenforms of slope strictly between 0 and 1, and $p^{-k}U^k_{p^2}(H_{\geq 1})$ is bounded. It suffices to show that $H$ is zero.

4.6.3. *The Shimura correspondence.* Suppose that $H \neq 0$. It follows that there exists, in $S^\dagger_{1/2}$, an eigenform of slope between 0 and 1. By the overconvergent Shimura correspondence [Ram09], there exists a corresponding overconvergent eigenform of weight $-2$, also with slope between 0 and 1. A priori, one might expect the level to be $\Gamma_0(288)$. However, after twisting, this form lies in $\Gamma_0(6)$, a fact that requires proof but follows from the underlying symmetry of $\eta$. Indeed:

4.6.4. **Lemma.** *The image of the Shimura correspondence from $S_{-1/2}(1, r)$ lies in*

$$S^\dagger_{-2}(\Gamma_0(6), r) \otimes \chi_{12},$$

*where $\chi_{12}$ is the quadratic character of conductor 12. Moreover, the eigenvalues of the corresponding form of level $\Gamma_0(6)$ of $U_2$ and $U_3$ are given by $2^{-2}$ and $3^{-2}$ respectively.*

One can provide fairly soft proofs of these type of facts using the trace formula — one only needs to compute that the appropriate spaces have the same traces of $(U_{p^2})^n$ and $(U_p)^n$ respectively.

From the existence of the eigencurve [CM98] (or by Coleman's results), it follows that there exists a classical form of weight $k \equiv -2 \mod p-1$ and slope between 0 and 1, as well as the indicated eigenvalues for $U_2$ and $U_3$. It follows from [BG09]

[6]Explicitly, choose an auxiliary prime $q$ distinct from 2, and then work at level $X(q)$, where the sheaf $\omega$ exists. Then, take invariants under $\mathrm{PGL}_2(\mathbf{F}_q)$.

Theorem 1.6 that if $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathcal{O})$ is the corresponding Galois representation, then

$$\overline{\rho}|D_p = \mathrm{Ind}_K^{\mathbf{Q}_p} \omega_2^{p-4} = \omega \otimes \mathrm{Ind}_K^{\mathbf{Q}_p} \omega_2^{-5},$$

and hence that

$$(\overline{\rho} \otimes \omega^4)|D_p = (\mathrm{Ind}_K^{\mathbf{Q}_p} \omega_2^{-5}) \otimes \omega^5 = (\mathrm{Ind}_K^{\mathbf{Q}_p} \omega_2^{-5p}) \otimes \omega_2^{5(p+1)} = (\mathrm{Ind}_K^{\mathbf{Q}_p} \omega_2^{5}).$$

Thus there exists a classical eigenform $h \in S_6(\Gamma_0(6), \mathcal{O})$ which is supersingular. Moreover, the eigenvalues of $U_2$ and $U_3$ are given by $2^4 \cdot 2^{-2} = 2^2$ and $3^4 \cdot 3^{-2} = 3^2$ respectively. On the other hand, $S_6(\Gamma_0(6), \mathcal{O})$ is one dimensional, and the corresponding eigenform

$$h = q + 4q^2 - 9q^3 + 16q^4 - 66q^5 - 36q^6 + \dots$$

has $U_2 = 4$ but $U_3 = -9$.

4.7. **An example: congruences for the partition function modulo powers of** 5**.** A routine computation shows that the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots$ of $U_{p^2}$ on $S^{\dagger}_{-1/2}(1, r)$ have slope

$$2, 7, 9, 15, 19, 22, 27, 29, 36, 39, \dots$$

It follows that one has an asymptotic expansion:

$$U_{p^2} \eta^{-1} \sim \alpha_1 \phi_1 + \alpha_2 \phi_2 + \alpha_3 \phi_3 + \dots$$

The fact that $\phi_1$ has slope 2 corresponds to the congruence for the partition function modulo powers of 5. In particular, it follows that:

$$\sum_{n=0}^{\infty} p\left(\frac{25^m n + 1}{24}\right) q^n \equiv \lambda_1^m \cdot \phi_1 + O(5^{7m}),$$

where $\phi_1$ is an eigenform for the Hecke operators $T_{\ell^2}$ with $\ell \neq 5$ as well as the operator $U_{25}$. Here one may numerically compute that

$$\lambda_1 = 4 \cdot 5^2 + 2 \cdot 5^3 + 3 \cdot 5^4 + 5^5 + 2 \cdot 5^7 + 3 \cdot 5^9 + 5^{11} + 5^{12} + 5^{14} + \dots$$

4.7.1. **Exercise** ($\star$)**.** *The eigenvalues of weight* $-2$ *of level* $N = 6$ *with* $p = 5$ *which are new at* 2 *and* 3 *with* $w_2 = w_3 = +1$ *have slope:*

$$v(\lambda_n) = v\left(5^{2n-1} \frac{(3n)!(3n)!}{(3n+1)!(3n-1)!} \frac{(6n+2)!(6n-2)!}{(2n)!(2n)!}\right)$$

4.7.2. **Exercise.** *One has the following congruences for* $c(n)$ *and* $p(n)$ *modulo other small primes:*

(1) *If* $n \equiv 0 \mod 2^a 3^b 5^c 7^d 11^e$ *and* $n \neq 0$*, then*

$$c(n) \equiv 2^{3a+8} 3^{2b+3} 5^{c+1} 7^d 11^e.$$

(2) *If* $24n \equiv 1 \mod 5^c 7^d 11^e$*, then*

$$p(n) \equiv 0 \mod 5^c 7^{[(d+2)/2]} 11^e.$$

*Explain these congruences in terms of the eigenvalue of* $U_p$ *or* $U_{p^2}$ *of smallest slope. Compute the slope of the next smallest eigenvalue in each case to give convergence results as above for* $p = 5$*.*

4.8. **An example: congruences for the partition function modulo powers of 5, following Watson.** Suppose instead of using Coleman's theory, one wanted to prove the congruence above directly, even just the considerably weaker classical congruences. Then one has to explicitly determine enough about the operator $U = U_{25}$ to show that (for example) it is divisible by 25. As in § 4.1, one needs to work explicitly with modular equations. For example, let

$$f(\tau) = 25\sqrt[4]{\frac{\Delta(5\tau)}{\Delta(\tau)}} = 25\frac{\eta(5\tau)^6}{\eta(\tau)^6} = 25q\prod_{n=1}^{\infty}\frac{(1-q^{5n})^6}{(1-q^n)^6},$$

4.8.1. **Lemma.** *The following identity holds:*

$$\prod_{m=0}^{4}\left(X - f\left(\frac{\tau+m}{5}\right)\right) = X^5 - a_1X^4 - a_2X^3 - a_3X^2 - a_4X - a_5,$$

*where*

$$\begin{aligned}
a_1 &= 5^2 f(63 + 260f + 315f^2 + 150f^3 + 25f^4)\\
a_2 &= 5^4 f(52 + 63f + 30f^2 + 5f^3)\\
a_3 &= 5^5 f(63 + 30f + 5f^2)\\
a_4 &= 5^7 f(6 + f)\\
a_5 &= 5^8 f
\end{aligned}$$

*Proof.* The proof is routine. □

4.8.2. **Lemma.** *For a non-negative integer $n$, let $A_n = \dfrac{U_5 f^n \eta^{-1}(\tau)}{\eta^{-1}(5\tau)}$ and $B_n = \dfrac{U_5 f^n \eta^{-1}(5\tau)}{\eta^{-1}(\tau)}$. Then $A_n$ and $B_n$ are polynomials in $f$ which satisfy the recurrence relation*

$$X_n = a_1X_{n-1} + a_2X_{n-2} + \ldots + a_5X_{n-5}.$$

*Moreover, for small values of $n$, we have the following equalities:*

$$\begin{aligned}
A_0 &= \frac{f}{5}\\
A_1 &= 5f(28 + 245f + 525f^2 + 455f^3 + 175f^4 + 25f^5)\\
A_2 &= 5^2 f(104 + 9100f + 113880f^2 + 528125f^3 + 1232725f^4 + 1660750f^5 + 1376375f^6\\
&\quad + 715000f^7 + 227500f^8 + 40625f^9 + 3125f^{10})\\
A_3 &= 5^3 f(19 + 13889f + 672885f^2 + 9791080f^3 + 66083900f^4 + 252487675f^5 + 608947625f^6\\
&\quad + 988926250f^7 + 1124158750f^8 + 913721875f^9 + 534909375f^{10} + 224081250f^{11} + 65609375f^{12}\\
&\quad + 12765625f^{13} + 1484375f^{14} + 78125f^{15})\\
A_4 &= 5^6 f(1 + 8375f + 1375975f^2 + 52547625f^3 + 831122125f^4 + 7023871875f^5 + 36454450625f^6\\
&\quad + 126528231250f^7 + 310499593750f^8 + 559393046875f^9 + 759056634375f^{10} + 788952734375f^{11}\\
&\quad + 634365468750f^{12} + 396053515625f^{13} + 191527734375f^{14} + 71064453125f^{15} + 19855468750f^{16}\\
&\quad + 4042968750f^{17} + 566406250f^{18} + 48828125f^{19} + 1953125f^{20})
\end{aligned}$$

$$B_0 = 1$$
$$B_1 = 5f(63 + 260f + 315f^2 + 150f^3 + 25f^4)$$
$$B_2 = 5^3 f(104 + 4095f + 32820f^2 + 107300f^3 + 182700f^4 + 180375f^5 + 107500f^6 + 38250f^7 + 7500f^8 + 625f^9)$$
$$B_3 = 5^4 f(189 + 49230f + 1512585f^2 + 15998850f^3 + 83171925f^4 + 251923750f^5 + 488490750f^6 + 640687500f^7 + 586327500f^8 + 379518750f^9 + 173362500f^{10} + 54750000f^{11} + 11390625f^{12} + 1406250f^{13} + 78125f^{14})$$
$$B_4 = 5^6 f(24 + 42920f + 4266360f^2 + 118018875f^3 + 1455608800f^4 + 9969720300f^5 + 42885018000f^6 + 125026746500f^7 + 259678080000f^8 + 397294462500f^9 + 457754050000f^{10} + 402607546875f^{11} + 272038500000f^{12} + 141147812500f^{13} + 55788750000f^{14} + 16505156250f^{15} + 3540625000f^{16} + 520312500f^{17} + 46875000f^{18} + 1953125f^{19}$$
$$\sum_{n=0}^{\infty} A_n T^n = \frac{q/5 + 5(28f + 182f^2 + 265f^3 + 140f^4 + 25f^5)T - 5^2(-104f - 20f^2 + 10f^3)T^2 - 5^4(-19f + 6f^2 + 5f^3)T^3 + 5^6 fT^4}{1 - a_1 T - a_2 T^2 - a_3 T^3 - a_4 T^4 - a_5 T^5}$$
$$\sum_{n=0}^{\infty} B_n T^n = \frac{1 - 20f(63 + 260f + 315f^2 + 150f^3 + 25f^4)T - 5^3 f(156 + 189f + 90f^2 + 15f^3)T^2 - 5^4 f(126 + 60f + 10f^2)T^3 5^6 f(6 + f)T^4}{1 - a_1 T - a_2 T^2 - a_3 T^3 - a_4 T^4 - a_5 T^5}$$

*Proof.* These follow from the standard methods. Note that some of these identities are quite classical, for example, $A_1 = f/5$ is just the identity

$$\sum p(5n+4)q^n = 5\frac{(1-q^5)^5(1-q^{10})^5(1-q^{15})^5\cdots}{(1-q)^6(1-q^2)^6(1-q^2)^6\cdots}$$

□

Many of these equations are (in slightly disguised form) in [Wat38]. These recurrences give enough information to prove (as in § 4.1) that, with respect to some suitable basis, that $U_{25} = 5^2\mathcal{F}$ for some suitable continuous operator $\mathcal{F}$, which allows one to prove the desired congruences, which is what Watson does. However, it does not seem obvious how one can use this approach to understand the *second* eigenvalue (and eigenvector) of $U_{25}$. What one needs to show is that the action on $U_{25}$ on some natural space is divisible by $5^7$ — yet this is only possible if one can somehow project away from the eigenform $\phi_1$ of slope 2. Unlike the Eisenstein series — which in weight 0 is just the constant 1 — there is no obvious way to account for the influence of $\phi_1$ when trying to estimate the error term.

## 5. $p$-adic arithmetic quantum chaos

(See [Sar95].)

Fix a modular curve $X = X(\Gamma)$. Don Blasius suggested to me the possibility that there could be an useful analogy between the discrete spectrum of the hyperbolic Laplacian

$$\Delta = -y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$$

and the operator $U_p$ in weight zero. In this section, we discuss various conjectures relating to this question.

One natural question is how to count such eigenvalues. They both form infinite countable sets — in the Archimedean case by a non-trivial result of Selberg, and in the $p$-adic case by Coleman — it suffices to note that the coefficients of the Fredholm determinant are all non-zero, which follows by a theorem of Koike [Koi75, Koi76].

The natural way to count eigenforms is thus by bounding the size of the eigenvalues. In the classical case, the classical result is the following:

5.0.1. **Theorem** (Weyl's Law)**.** *Let* $N(T) := \#\{\lambda \mid v(\lambda) < T\}$ *denote the counting function for discrete eigenvalues of* $\Delta$*. Then*

$$N(T) \sim \frac{\mathrm{Vol}(X)}{4\pi} \cdot T.$$

The modular curves inherit from $\mathbf{H}$ a natural metric of constant curvature -1. Recall that, with respect to this metric,

$$\mathrm{Vol}(X) = [\Gamma_0(1) : \Gamma] \cdot \frac{\pi}{3}.$$

Somewhat better error bounds are known, but they, too, are non-trivial. Note that volume is taken with respect to the natural measure on $X$ which makes it a manifold of constant curvature $-1$. When it comes to counting finite slope eigenforms, the natural measure of size of eigenvalues is by their valuation. Since the operator $U_p$ depends on the choice of a subgroup of order $p$, it is more natural to work with $X_0(p)$ rather than $X$ (at the level of the appropriate rigid analytic spaces $X^{\mathrm{rig}}[r]$ there is no difference due to the existence of the canonical subgroup).

5.0.2. **Conjecture** ($p$-adic Weyl's Law)**.** *Let* $N(T) := \#\{\lambda \mid v(\lambda) < T\}$ *denote the counting function for eigenvalues of* $U_p$*. Then*

$$N(T) \sim \frac{\mathrm{Vol}(X_0(p))}{4\pi} \cdot T.$$

How does one count such eigenvalues? The slopes of the eigenvalues are determined (via Newton's Lemma) to the valuations of the Fredholm power series of $1 - TU$. We have the following partial result, which proves one direction of Conjecture 5.0.2:

5.0.3. **Theorem.** *There is an inequality:*

$$N(T) \leq \frac{\mathrm{Vol}(X_0(p))}{4\pi} \cdot T + o(T).$$

This follows from the estimates of Wan [Wan98] (in particular, it follows via an easy computation from Lemma 3.1 of *ibid.*) One also obtains from this a bound:

5.0.4. **Theorem.** *There is an inequality:*

$$N(T) \gg \frac{\mathrm{Vol}(X_0(p))}{4\pi} \cdot T,$$

*where the implicit constant depends on* $X$*.*

Another natural problem to be concerned with is the behavior of the eigenfunctions $U\phi = \lambda\phi$ as *functions* as $\lambda \to \infty$. For example:

(1) How fast does the $L_\infty$-norm grow with $\lambda$? Obviously this depends on some normalization of the eigenfunctions $\phi$. Since the eigenfunctions are $L^2$, a natural normalization is the $L^2$-norm, i.e., insisting that

$$\|\phi\|_2^2 = |\langle \phi, \phi \rangle| = 1.$$

(2) What is the distribution of the zeroes of $\phi$? More generally, what do the eigenfunctions $\phi$ look like as functions?

For modular surfaces (at least in the compact case, although a lot is known in the open case as well), the answers to these questions are as follows:

(1) There is a general bound for surfaces ([SS89])

$$\|\phi\|_\infty = O(\lambda^{1/4}).$$

In the specific case of $\Delta$ on compact *arithmetic* surfaces, one has

$$\|\phi\|_\infty = O(\lambda^{5/32})$$

by Iwaniec–Sarnak [IS95], who moreover conjecture that

$$\|\phi\|_\infty =^? O(\lambda^\epsilon).$$

Iwaniec–Sarnak (*ibid.*) also prove the lower bound

$$\|\phi\|_\infty \gg \log\log\lambda$$

holds for infinitely many $\phi$.

(2) Roughly speaking, the $\phi$ become "equidistributed" over $X$ as $\lambda$ becomes arbitrarily big, and the point measure based on the zeros (of eigenforms) converges to the point measure.

5.0.5. **Remark.** In the non-Archimedean case, there is an issue concerning how to take normalizations. One natural normalization is given by $q$-expansions, namely, assuming that the leading coefficient of $\phi$ is $q$. A different possible normalization is given by insisting that

$$|\langle\phi,\phi\rangle| = 1,$$

as in the arithmetic case. One issue with this is that *it is not even clear that this normalization is possible.* The point is that one does not know, given $\lambda \neq 0$, whether the generalized $\lambda$-eigenspace of $U$ is a *genuine* eigenspace. This is because the pairing does not give rise to a Hilbert space structure which doesn't seem to exist in the non-Archimedean world. For example, suppose that $(U-\lambda)\psi = \phi$ and $(U-\lambda)\phi = 0$. Then

$$\langle\phi,\phi\rangle = \langle(U-\lambda)\psi,\phi\rangle = \langle U\psi,\phi\rangle - \lambda\langle\psi,\phi\rangle = \langle\psi,U\phi\rangle - \lambda\langle\psi,\phi\rangle = \langle\psi,\lambda\phi\rangle - \lambda\langle\psi,\phi\rangle = 0.$$

The semi-simplicity is still unknown even in the classical case, see Remark 3.8.9.

We do, at least, have the following estimate:

5.0.6. **Lemma** (Cauchy–Schwartz)**.** *Let $\alpha$, $\beta \in M_0^\dagger(\Gamma, r)$, with $r = 1/2$. Let $\|\cdot\| = \|\cdot\|_{1/2}$. There is an inequality:*

$$|\langle\alpha,\beta\rangle| \leq \|\alpha\|\|\beta\|.$$

In particular, one could "define" $\|\phi\|_\infty$ to be the quantity

$$\|\phi\|_\infty := \frac{\|\phi\|}{\sqrt{|\langle\phi,\phi\rangle|}},$$

then $\|\phi\|_\infty$ does not depend on $\phi$ up to scalar, and is conjecturally finite for eigenforms, but may (and will) be infinite in general.

5.0.7. **Exercise.** *Prove that given any two functions $\phi$ and $\psi$, there exists some non-trivial linear combination $\alpha\phi + \beta\psi$ such that $\langle\alpha\phi+\beta\psi, \alpha\phi+\beta\psi\rangle = 0$.*

We fix a radius of convergence $r$, and let $\|\cdot\|$ denote the supremum norm. In this case, we have:

(1) We view $\|\phi\| = \|\phi\|_r$ as a substitute for the $L^\infty$-norm (it is a supremum norm).
   (a) There is no known upper bound for $\|\phi\|$.
   (b) There is a trivial lower bound $\|\phi\| \geq 1$, if we normalize by using $q$-expansions, but no known non-trivial bounds.
   (c) If $r \leq 1/2$, there is a trivial lower bound $\|\phi\| \geq 1$, if we normalize by setting
$$|\langle \phi, \phi \rangle| = 1.$$
(2) Regarding the function $\phi$ for large $\lambda$, there are two natural questions one could ask depending on the normalization.
   (a) On the ordinary locus, the functions $\phi$ can be thought of as elements of the universal deformation ring of a finite number of residual representations. Nothing is known about the distribution of these points.
   (b) On the supersingular locus, a result of Buzzard [Buz03] implies that eigenforms $\phi$ extend to sections of $X^{\mathrm{rig}}[r]$ for *all* $r < 1$, and that they cannot be extended beyond this (so $\|\phi\|_r \to \infty$ as $r \to 1$), but nothing is known concerning what these functions look like.

5.0.8. **Lemma** (Hadamard three-circle theorem)**.** *Suppose that $\|F\|_a = p^A$ and $\|F\|_b = p^B$ for rational $0 < a < b < p/(p+1)$. Then there is an inequality:*
$$\log_p \|F\|_r \leq A + (B - A)\frac{(r - A)}{(b - a)}$$
*for all $a \leq r \leq b$ in $\mathbf{Q}$. If $F$ has no zeroes on the corresponding annulus, then equality holds.*

*Proof.* Both $A$ and $B$ are rational. Since norm of an integral power of $F$ is the corresponding power of the norm, after replacing $F$ by a power of itself and multiplying the result by a power of $p$, we may assume that $A = ma$ and $B = mb$ for some $m \in \mathbf{Z}$. Now consider the function:
$$G = F \cdot t^{-m}.$$
By construction, the norm of $G$ on the annulus $|t| = |p^a|$ is 1, and the norm on the annulus $|t| = |p^b|$ is also 1. If $F$ has no zeroes, the same argument applies to $F^{-1}$. □

One consequence is that the minimum value of $\|F\|_s \|F\|_{1-s}$ for $s \in (1 - r, r)$ and $F \in M_0^\dagger(\Gamma, r)$ with $r > 1/2$ occurs for $s = 1/2$.

5.1. **An explicit example: $N = 1$ and $p = 2$.** When $N = 1$ and $p = 2$, some mileage[7] may be obtained from the fact that $X_0(2)$ has genus zero, as well as the fact that $U_2$ has such an explicit form on $M_0^\dagger(1, r)$, namely by identifying the latter with the Tate algebra
$$M_0^\dagger(1, r) \simeq \mathbf{C}_2\langle 2^r \cdot f \rangle, \qquad f = q\prod_{n=1}^{\infty}(1 + q^n)^{24}.$$

[7] or kilometrage, if you prefer.

With respect to the natural basis in terms of powers of $2^r \cdot f$, one has $U_2 = [s_{ij}]$, where

$$s_{ij} = \frac{3i(i+j-1)!2^{2i+2j-1}}{(2i-j)!(2j-i)!} \cdot 2^{(6-12r)(i-j)}.$$

When $r = 1/2$ this is particularly symmetric. We will (mostly) be concerned with this value of $r$, although not exclusively. Note that one has the following relationship between the valuation of $f$ and the annuli $|t| = |2^r|$, which can be deduced in a similar manner to the computations in §3.2.2:

5.1.1. **Lemma.** *Suppose that $0 < r < 1$. Then, on the annuli $|t| = |2^r|$, one has $|f| = \|f\|_r = 2^{12r}$.*

For convenience, however, we make the following definitions.

5.1.2. **Definition.** *Let $\|\cdot\|$ denote the norm $\|\cdot\|_{1/2}$, and let $g = 2^6 f$.*

Note that $\|g\| = 1$. Let us denote the eigenvectors by $\phi_n$ for positive integers $n$. A key result of [Buz03] implies that eigenvectors $\phi$ can be analytically continued to be sections of $X^{\mathrm{rig}}[r]$ for any $r < 1$. One has an exact formula for the slopes of the eigenvalues [BC05], and one knows the spectral conjecture [Loe07]. For example, the slope of the $n$th eigenvalue $\lambda_n$ of the eigenfunction $\phi_n$ is

$$v(\lambda_n) = 1 + 2v\left(\frac{(3n)!}{(n)!}\right).$$

5.1.3. **Exercise.** *Using the explicit formulae for the slopes, prove the 2-adic Weyl's law for $N = 1$, namely, that*

$$N(T) \sim \frac{\mathrm{Vol}(X_0(2))}{4\pi} \cdot T + O(\log(T)).$$

Recall the matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ and the diagonal matrix $[D_{ii}]$ from [BC05], where:

$$a_{ij} = 2^{(j-i)(12-6r)} 6ij \left(\frac{(2j)!}{2^j j!}\right)^2 \left(\frac{2^i i!}{(2i)!}\right)^2 \left(\frac{(2i-1)!}{(i+j)!}\right) \left(\frac{(2j+i-1)!}{(3j)!}\right) \binom{j}{i-j},$$

$$b_{ij} = \frac{j}{i} a_{ji},$$

$$d_{ii} = \frac{2^{4i+1}(3i)!^2 i!^2}{3(2i)!^4}.$$

One has a factorization $U = ADB$. When $r = 1/2$, the matrices $A$ and $B$ lie in $\mathbf{Z}_2$ and are congruent to the identity modulo 2. As noted in [Loe07], the same holds for all $r$ in the range $5 < 6r < 7$. In particular, if we let $\alpha = 2^{(6r-12)}$, then the proof of the main theorem of Loeffler [Loe07] implies the following:

5.1.4. **Lemma.** *The $n$-th eigenfunction $\phi_n$, considered as an element of the Tate algebra $\mathbf{C}_2\langle \alpha g \rangle$, admits, after normalization, an expansion:*

$$\phi_n = \sum a_i (\alpha g)^i$$

*where $v(a_i) > 0$ for all $i \neq n$, and $v(a_n) = 0$.*

We derive some immediate corollaries from this.

5.1.5. **Corollary.** *$\phi_n$ has exactly $n$ zeroes in $X^{\mathrm{rig}}[r]$ for any $5/12 < r < 7/12$. In particular, any eigenform $\phi_n$ does not vanish on the annuli $|t| = |p^r|$ for $r$ in that range.*

*Proof.* This follows from the Weierstrass preparation theorem. Note that exactly one of the zeroes occurs in the ordinary locus (at the cusp $\infty$). □

As an example, consider the point $\tau = i$, one has $j = 1728$ and $f = -2^{-6}$ or $g = -1$, which lies on the annulus $|t| = 1/2$. Hence, no eigenform vanishes at (or anywhere near) this point. On the other hand, an easy exercise shows that every classical modular form of level $4^m$ vanishes at $i$.

5.1.6. **Exercise** ($\star$)**.** *Where are the zeros for large $n$?*

5.1.7. **Corollary.** *Suppose that $n$ is odd. Then*

$$|\langle \phi_n, \phi_n \rangle| = \|\phi_n\|^2.$$

*Proof.* Writing $\phi = \sum a_i g^i$ with $a_i \in \mathbf{Z}_2$, and $\phi \equiv g^n \mod 2$. we have

$$\langle \phi_n, \phi_n \rangle = \sum a_i a_j \langle g^i, g^j \rangle = \sum k a_k^2 \equiv n a_n \mod 2,$$

which, if $n$ is odd, has valuation 1. On the other hand,

$$\|\phi_n\| = \sup |a_i| \|g^i\| = \sup |a_i| = 1,$$

since $\|g\| = 1$. □

5.1.8. **Corollary.** *Normalize the $\phi_n$ so that $\|\phi_n\| = 1$. Then the limit:*

$$\lim_{\to} \|\phi_n\|$$

*exists as a continuous $\mathbf{R}$-valued function on $X^{\mathrm{rig}}[1/2]$, and coincides with the locally constant function which is 1 on the annulus $|t| = |p^{1/2}|$ and zero everywhere else.*

There also exists a second natural normalization of the eigenforms $\phi$, namely, the one given by $q$-expansions. In particular, for these normalizations, we can take the $q$-expansion norm $\|\phi\|_0$. Since the $\phi$ are eigenforms, the $q$-expansion norm can be read off from the coefficient of $g$ in $\phi$, e.g., if $\phi = \sum a_i g^i$, then $\|\phi\|_0 = |2^6 \cdot a_1|$.

5.1.9. **Lemma.** *There is a lower bound:*

$$\frac{\|\phi_n\|}{\|\phi_n\|_0} \geq 2^{n+5}.$$

*Proof.* Let us write $\phi_n = \sum a_i g^i$ with $a_i \in \mathbf{Z}_2$ and $\phi_n \equiv g^n \mod 2$. Then we may also write:

$$\alpha^n \phi_n = \sum \alpha^{n-i} a_i (\alpha g)^i,$$

and by Lemma 5.1.4 we deduce that:

$$(n-i)v(\alpha) + v(a_i) > 0$$

for all $-1 < v(\alpha) < 1$. With this normalization, we have $\|\phi_n\| = 1$, and $\|\phi_n\|_0 = |2^6 \cdot a_1|$. Yet, taking $v(\alpha) \to -1$, we deduce that $v(a_1) \geq n-1$, and hence $v(2^6 \cdot a_1) \geq n+5$. The result follows. □

5.1.10. **Guess.** *Suppose that* $\phi_n \equiv g^n \mod 2$. *Then*

$$|\langle \phi_n, \phi_n \rangle| = |\langle g^n, g^n \rangle| = |n|.$$

*In particular,*

$$\|\phi_n\|_\infty^2 := \frac{\|\phi\|^2}{|\langle \phi, \phi \rangle|} = \frac{1}{|n|} = O(\log v(\lambda)).$$

*Moreover, there are equalities:*

$$\frac{\|\phi_n\|_0}{|\langle \phi_n, \phi_n \rangle|} = |2^9 n \cdot \lambda_n|, \qquad \frac{\|\phi_n\|_0^2}{\|\phi_n\|^2} = |2^9 n^2 \cdot \lambda_n|.$$

I must admit the first equality is based on an embarrassingly small amount of data (for $n = 1$ to 4), although the final identity is similar to one guessed by Loeffler when $N = 1$ and $p = 5$. If we compare the conjectural lower bound of $\|\phi_n\|/\|\phi_n\|_0$ to the bound established in Lemma 5.1.9 (to check for consistency), we obtain the estimate:

$$v(\lambda_n) + 2v(n) \geq 2n + 1.$$

This is easy to prove directly (given the explicit formula for the slopes in this case), and equality holds only for $n = 1$. From Weyl's Law, we actually have $v(\lambda_n) \sim 4n$.

5.1.11. **Guess.** *The zeros of* $\phi_1$ *occur when* $v(A(E)) = r$ *takes the following values: once when* $r = 8$, *and then* $2^n$ *times for integers* $n \geq 1$, *when*

$$r = \frac{1}{12}\left(12 - \frac{2}{2^n}\right).$$

*The zeroes of* $\phi_2$ *occur when* $v(A(E)) = r$ *takes the following values: once when* $r = 3$, *and then* $2^n$ *times for integers* $n \geq 1$, *when*

$$r = \frac{1}{12}\left(12 - \frac{4}{2^n}\right).$$

This guess is equivalent to the following. Write $\phi_1 = \sum a(n)h^n$, then the Newton polygon of this power series occurs at the vertices $(1, 0)$, $(2, 8)$, and

$$(2^n + 1, 24 \cdot 2^n - 2n - 6).$$

5.1.12. **Exercise.** *Verify this for* $r < 1 - \epsilon$ *for some small* $\epsilon$.

5.2. **Overconvergent $p$-adic arithmetic quantum unique ergodicity.** In the spirit that the section heading suggests[8], we make the following general guesses:

5.2.1. **Guess.** *Consider the space* $M_0^\dagger(\Gamma, 1/2)$. *Then the following hold.*

(1) **The Spectral Conjecture**: The operator $U_p$ admits a convergent spectral expansion, and the action of $U_p$ is semi-simple.
(2) **$p$-adic Weyl's law**: The number of eigenvalues of slope at most $T$ satisfies Weyl's law.
(3) If $\|\cdot\|$ denotes the supremum norm, then

$$1 \leq \frac{\|\phi\|^2}{|\langle \phi, \phi \rangle|} = O(\log v(\lambda)).$$

[8]Thanks to Simon Marshall for the satisfying acronym.

(4) If $\|\cdot\|_0$ denotes the $q$-expansion norm, then

$$\frac{\|\phi\|^2}{|\langle\phi,\phi\rangle|\cdot|\lambda|} = O(\log v(\lambda)).$$

(5) As $\lambda \to 0$, are the functions $\phi$ are distributed in some natural way? As a special case, is it true that the (normalized) sequence

$$\|\phi\|^2$$

on $M_0^\dagger(\Gamma, 1/2)$ converges to functions which are constant on supersingular annuli $|t| = 1/2$, at least if one restricts to subsequences for which the residual representation $\overline{\rho}_\phi$ is constant? (Or, perhaps, to a connected component of the eigencurve?) Do the sum of delta measures on $X^{\mathrm{rig}}[1/2]$ supported on the zeros of $\phi$ converge to any explicit measure on the corresponding Berkovich space (already an interesting computation for $N = 1$ and $p = 2$)? To make an even wilder guess, let $S_\phi$ denote the zero set of $\phi$ in the region $X^{\mathrm{rig}}[1/2]$. We expect (and know for $N = 1$ and $p = 2$) that $|S_\phi|$ grows linearly with respect to the natural ordering of the eigenvalues. It may also be the case that $S_\phi$ is completely contained within $X^{\mathrm{rig}}[r]$, where $r = 1/(p+1)$. Consider the measures:

$$\frac{1}{|S_\phi|}\sum_{x\in S_\phi}\delta_x.$$

Is it the case that these measures on complex valued continuous functions on $X^{\mathrm{rig}}[r]$ have a limiting measure on the Berkovich space associated to the affinoid $X^{\mathrm{rig}}[r]$ (for $r = 1/(p+1)$)? If so, does it converge to the Gauss point corresponding to the supremum norm on the entire space? For example, is it the case that when $N = 1$ and $p = 2$, and for a polynomial $F \in \mathbf{C}_2[[g]]$, one has

$$\lim_{\lambda\to 0}\frac{1}{|S_\phi|}\sum_{x\in S_\phi}|F(x)| = \|F(x)\|_r,$$

where $r = 1/(p+1)$. For example, if $F = g$, this is equivalent to saying that almost all of the zeros of $\phi$ (in $X^{\mathrm{rig}}[1/2]$), the Hasse invariant has valuation at least $r - \epsilon$ for any fixed $\epsilon > 0$ and $r = 1/(p+1)$.

(6) As $\lambda \to 0$, the Galois representations $\rho_\phi$ are distributed on the corresponding global deformation rings $\mathrm{Spec}(R_{\overline{\rho}})$ with respect to a natural measure. Note that when $N = 1$ and $p = 2$ all the eigenfunctions have coefficients in $\mathbf{Z}_2$; In general, Buzzard raises the question [Buz05] of whether for any $N$ and $p$ *all* finite slope eigenforms in any particular weight are defined over a fixed extension $K/\mathbf{Q}_p$. Hence, by measure, we are considering subsets of the compact $p$-adic manifold $\mathrm{Hom}(R_{\overline{\rho}}, \mathcal{O}_K)$ rather than some measure on the Berkovich space associated the rigid analytic space corresponding to $R_{\overline{\rho}}$.

Assuming a very strong version of the Gouvêa–Mazur conjecture, one can rephrase part 6 of this guess as follows (and equally vaguely):

5.2.2. **Guess.** *Consider the classical modular eigenforms of weight $p^{k-1}(p-1)$ over $\overline{\mathbf{Q}}_p$. Then, as $k \to \infty$, the Galois representations modulo $p^k$ are distributed on the corresponding global deformation rings with respect to a natural measure.*

On the other hand, the claims concerning the distribution of $\phi$ on the annuli are close to meaningless without some possible candidate distribution.

## 6. Student projects

There are various projects, depending on the inclination of the student — some are more theoretical and some are more computational. (Of course, the computations should help with the theoretical musings.)

6.1. **Turn Guess 5.2.1 into a conjecture.** or at least a question. This requires:

6.1.1. *More data.* Suppose that $N = 1$ and $p = 2$. Here's a practical way of computing eigenforms of high slope. Choose an arbitrary cut off, say $n = 100$; let $M = [s_{ij}]$ for $i, j \leq 100$ denote the corresponding matrix. Compute the characteristic polynomial $X^{100} + \ldots$ of $M$. The roots of this polynomial all lie in $\mathbf{Z}_2$, so they are easy to compute to high 2-adic accuracy. Let $\lambda'$ denote a root of this polynomial to high 2-adic accuracy. Let $\lambda$ denote the corresponding genuine eigenvalue. Choose a a random vector $u \in \mathbf{Z}_2^{100}$, and let

$$v = (M - \lambda')^{-m} u.$$

for some largish integer $m$. Then $v$ should be a good approximation to the genuine eigenvector associated to $\lambda$. For example, the eight eigenvalue $\lambda_8$ has valuation 31, and, using Hensel's Lemma, we compute that

$$\lambda_8 = 18020903046061192281127374673614608115989037626 0$$
$$12182154057384464387035523314270868146107543715 84 + O(2^{321})$$

Let $\lambda'$ denote this number, and let $M$ be the $50 \times 50$ matrix $[s_{ij}]$ with $i, j = 1, \ldots, 50$. Let

$$v = (M - \lambda')^{-100}(1, 0, 0, 0, \ldots, 0).$$

(A larger exponent would probably give a more accurate approximation, but I didn't do this in a very clever way so even this computation was a little slow.) Let $w$ denote the scalar multiple of $v$ normalized so that the first entry is $2^{-6}$. An approximation to $\phi_8$ should then be given by

$$\phi_8 \sim \sum_{i=1}^{50} v_i g^i.$$

We compute the valuations of the coefficients $v_i$ to be as follows:

$$[-6, -9, -9, -14, -13, -16, -16, -23, -16, -16, -13, -14,$$
$$-9, -9, -6, -7, 4, 6, 11, 12, 19 \ldots]$$

To test this as an approximation to $\phi_8$, note that the square of the norm $\|\phi_8\|^2$ appears to be equal to $2^{46} = 2^9 \cdot 8^2 \cdot 2^{23}$, as predicted by Guess 5.1.10. We also compute the first few terms of the $q$-expansion (omitted, because they are ratios of 500000 digit numbers, although it is ridiculous to compute them in this manner, since one should work modulo some power of 2 — hopefully some of you can programme better than I can):

$$\phi_8 \simeq q + a(2)q^2 + a(3)q^3 + a(4)q^4 + a(5)q^5 + a(6)q^6 + \ldots$$

we find that $v(a(2)) = 31 = v(\lambda_8)$, and we also check that

$$a(3)a(5) \equiv a(15) \mod 2^{115},$$

which is a good check that this is actually an eigenform. We can compute that the zeroes of $\phi_8$ in $X^{\mathrm{rig}}[1/2]$ occur at the cusp $q = 0$ and on $|t| = r$ for

$$r = \{1/4, 7/24, 7/24, 5/16, 5/16, 5/16, 5/16\}.$$

Do there exist $\phi$ with zeros on the annulus $|t| = |p^r|$ with $1/3 < r < 2/3$? Do almost all of the zeroes of $\phi$ with $r \leq 1/2$ have valuation $1/3 - \epsilon$? A possibly dodgy computation for $\phi_{12}$ found zeroes of the following valuations (away from $q = 0$):

$$r = \{1/4, 7/24, 7/24, 5/16, 5/16, 5/16, 5/16, 1/3, 1/3, 1/3, 1/3\}.$$

A computation for $\phi_{64}$ (no attempt to be careful about accuracy) with $\lambda_{64} = 2^{255} + \ldots$ yields $\phi_{64} \sim^? \sum a_i g^i$ with the following valuations for the coefficients $a_i$ (normalized so that $a_1 = -6$):

$$\begin{aligned}[-6, -9, -9, -14, -13, -16, -16, -23, -21, -24, -24, -29, -28, -31, -31, -40, -37, -40, -40,\\ -45, -44, -47, -47, -54, -52, -55, -55, -60, -59, -62, -62, -73, -69, -72, -72, -77, -76,\\ -79, -79, -86, -84, -87, -87, -92, -91, -94, -94, -103, -100, -103, -103, -108, -107, -110,\\ -110, -117, -115, -118, -118, -123, -122, -125, -125, -138, -125, -125, -122, -123, -118,\\ -118, -115, -117, -110, -110, -107, -108, -103, -103, -100, -103, -94, -94, -91, -92, -87,\\ -87, -84, -86, -79, -79, -76, -77, -72, -72, -69, -73, -62, -62, -59, -60 \ldots]\end{aligned}$$

which yields $\|\phi_{64}\| = 2^9 \cdot 64^2 \cdot 2^{255} = 2^{276} = 2^{2 \cdot 138}$, and has roots with $|t| = |p^r|$ and $r \leq 1/2$ with $r$ as follows:

$$\begin{aligned}r = \{&1/4, 7/24, 7/24, 5/16, 5/16, 5/16, 5/16, 31/96, 31/96, 31/96, 31/96, 31/96, 31/96, 31/96,\\ &31/96, 21/64, 21/64, 21/64, 21/64, 21/64, 21/64, 21/64, 21/64, 21/64, 21/64, 21/64, 21/64,\\ &21/64, 21/64, 21/64, 21/64, 127/384, 127/384, 127/384, 127/384, 127/384, 127/384, 127/384,\\ &127/384, 127/384, 127/384, 127/384, 127/384, 127/384, 127/384, 127/384, 127/384, 127/384,\\ &127/384, 127/384, 127/384, 127/384, 127/384, 127/384, 127/384, 127/384, 127/384, 127/384,\\ &127/384, 127/384, 127/384, 127/384, 127/384\},\end{aligned}$$

Note that one has

$$\frac{4}{|S_{\phi_{64}}|} \sum_{x \in S_\phi} |g(x)| = 0.92577 \ldots \simeq 4\|g\|_{1/3} = 1,$$

which (if correct) might be taken as some sort of evidence. In comparison, the sets $S_n$ of $n$th roots of unity also converge to the Gauss norm on the affinoid corresponding to the closed unit disk, and one has

$$\frac{1}{|S_{64}|} \sum_{\zeta^{64}=1} |\zeta - 1| = \frac{1}{64} \sum_{n=0}^{5} 2^n \cdot 2^{-1/2^n} = 0.92577 \ldots \sim \|z - 1\| = 1.$$

(The numerical coincidence is not accidental — the LHS are literally equal, which is perhaps surprising but not completely preposterous since both are sums of powers of two with exponents in $\frac{1}{32}\mathbf{Z}$ — if $|t(x)| = |2^r|$ with $r = 127/384$, for example, then $|g(x)| = 2^{-6}2^{12r} = 2^{-65/32}$, and so $4|g(x)| = 2^{-1/32}$. In comparison, the roots of $\phi_{63}$ all occur with $r = 1/3$ with the exception of the cusp $\infty$ and one root for $r = 1/4$, so

$$\frac{4}{|S_{\phi_{63}}|} \sum_{x \in S_\phi} |g(x)| = \frac{41}{42} = 0.97619 \ldots \simeq 4\|g\|_{1/3} = 1.$$

6.2. **More precise questions.** What does it *mean* for the eigenvectors $\phi_j$ themselves to become "equidistributed"? Compute lots of eigenfunctions $\phi_j$ for $N = 1$ and $p = 2$ and then stare at them, and think about $p$-adic equidistribution and Berkovich spaces.

6.3. **Some Guesses.** Let $N = 1$ and $p = 2$, and let $k \in \mathbf{Z}_2$. Let $j$ denote a sequence of positive integers tending to infinity such that $j \to k$ in $\mathbf{Z}_2$. Do the Galois (pseudo-)representations $\phi_j$ tend to a limit? For example, if $j \to 0$, does $\phi_j$ tend to $1 \oplus \epsilon^{-1}$ where $\epsilon$ is the cyclotomic character? Numerically, one (seems to have)

$$\phi_{64} \equiv q + \left(1 + \frac{1}{3}\right) q^3 + \left(1 + \frac{1}{5}\right) q^5 + \left(1 + \frac{1}{7}\right) q^7 + \ldots \mod 2^{19},$$

suggesting that $\phi_{2^m}$ converges to

$$\begin{aligned} \sum_{n=1}^{\infty} \left( \sum_{d|n}^{d \text{ odd}} \frac{1}{d} \right) q^n &= h - 24h^2 + \frac{2560}{3} h^3 - 35840h^4 + \ldots \\ &= \sum_{n=1}^{\infty} \binom{2n}{n} \frac{(-1)^{n-1}}{32n} (16h)^n \\ &= \frac{1}{16} \log \left( \frac{1 + \sqrt{1 + 64h}}{2} \right) \end{aligned}$$

Note (by inspection) that this function lies in $M_0^\dagger(1, r)$ for all $r < 1/3$, but *not* for $r = 1/3$ (this is also consistent with Lemma 3.8.5, because it lies in the kernel of $U_2$). Moreover, the zeroes of this function occur exactly when $16h = \zeta^2 - \zeta$ for a root of unity $\zeta$ with $|\zeta - 1| < 1$, equivalently, for a root of unity of two power order. This is also consistent with the computations above. Another reason one might gues this convergence is that $\phi_{2^n}$, which has slope $2^{n+2} - 1$, lives in a Coleman family — and if the radius of the family with constant slope is very large (exponential rather than linear as predicted by the Gouvêa–Mazur conjecture) then it will pass through the evil Eisenstein series of weight $2^{n+2}$.

6.4. **Trace formula methods.** Consider the question of how the Galois representations $\rho_j$ associated to $\phi_j$ are distributed. For $N = 1$ and $p = 2$, they all have coefficients in $\mathbf{Z}_2$, so they land in the $\mathbf{Q}_2$-points of the universal deformation ring of $\overline{\rho}$. (More accurately, there is only a universal pseudo-deformation ring, and a big Hecke algebra $\mathbf{T}$.) The corresponding big Hecke algebra is presumably a quotient of a power series ring over $\mathbf{Z}_2$ in a small number of variables. Probably those variables can be chosen to map to $T_l$ for small primes $l$. One may then study $T_l$ using the $p$-adic trace formula (for $p = 2$ here, but also more generally). Specifically, one may compute the trace of any *compact* operator on $M_k^\dagger(r)$. Hence suitable test functions are continuous maps composed with $U_p$, for example $T_l U_p$ for any prime $l$.

6.5. **Rigorous arguments.** Can one prove/improve any of the upper or lower bounds for $\|\phi_n\|$ or $|\langle \phi_n, \phi_n \rangle|$? Can one prove any useful bounds at all for general $N$ and $p$?

6.6. **The Spectral conjecture.** Can one prove anything? For example:

6.6.1. *$p$-adic Adjoint L-functions.* Prove Exercise 3.11.1.

6.6.2. *Symmetric Matrices.* What restrictions — if any — does the existence of the invariant pairing $\langle *, * \rangle$ put on the spectrum of $U$ on $M_0^\dagger(\Gamma, 1/2)$? For example: determine whether their exists an $\infty \times \infty$ matrix $M$ with coefficients in $\mathbf{Z}_p$ such that:

(1) $M$ is symmetric. ($m_{ij} = m_{ji}$.)
(2) If $B$ is the Banach space of convergent sequences in $\mathbf{Q}_p$, then $M$ acts compactly on $B$ in the natural way. ($\lim m_{ij} = 0$.)
(3) The kernel of $M$ is trivial.
(4) The characteristic power series of $M$ is trivial; equivalently, $M$ is topologically nilpotent; equivalently, the trace of $M^n$ is zero for all $n > 0$.

6.6.3. *Integral structures.* Can one find *canonical* integral structures on $M^\dagger(\Gamma, r)$ on which the action of $U$ is (close to) semi-simple on the mod-$p$ reduction? This is already interesting and difficult on the space of classical modular forms.

6.6.4. *The Slope conjectures.* Due to Buzzard [Buz05], Lisa Clay, and others. Buzzard's conjecture has associated `pari.gp/magma` scripts (see the paper); play around with those programs if you can.

6.6.5. *Applications to congruences.* How often does one expect there to be a form of slope $\mu$ with $0 < \mu < 1$ and weight 0?

6.7. **Some reading.** It might be worthwhile to take a look at the paper of Gouvêa and Mazur [GM95] — it's a very easy read. For the classical take on these congruences, look at Waton's paper [Wat38] (For a later, similar approach, see the paper by Atkin and O'Brien [AO67].) All the technical fact concerning modular forms we will need are mostly in the first chapter of Katz's Antwerp paper [Kat73]. Remind yourself what the spectral theorem for compact operators is. It might be useful to read [Loe07] and perhaps scan [BC05] for some computations with $N = 1$ and $p = 2$. Feel free to look at Coleman's papers, although note that we won't require the full machinery he uses (and develops) because we will be working in fixed weight.

## References


[AO67] A. O. L. Atkin and J. N. O'Brien, *Some properties of $p(n)$ and $c(n)$ modulo powers of 13*, Trans. Amer. Math. Soc. **126** (1967), 442–459. MR 0214540 (35 #5390)

[AS86] Avner Ash and Glenn Stevens, *Cohomology of arithmetic groups and congruences between systems of Hecke eigenvalues*, J. Reine Angew. Math. **365** (1986), 192–220. MR 826158 (87i:11069)

[BC05] Kevin Buzzard and Frank Calegari, *Slopes of overconvergent 2-adic modular forms*, Compos. Math. **141** (2005), no. 3, 591–604. MR 2135279 (2005k:11106)

[BC06] ———, *The 2-adic eigencurve is proper*, Doc. Math. (2006), no. Extra Vol., 211–232 (electronic). MR 2290588 (2007j:11055)

[BCD+08] Matthew Baker, Brian Conrad, Samit Dasgupta, Kiran S. Kedlaya, and Jeremy Teitelbaum, *p-adic geometry*, University Lecture Series, vol. 45, American Mathematical Society, Providence, RI, 2008, Lectures from the 10th Arizona Winter School held at the University of Arizona, Tucson, AZ, March 10–14, 2007, Edited by David Savitt and Dinesh S. Thakur. MR 2482343 (2010a:14001)

[BG09] Kevin Buzzard and Toby Gee, *Explicit reduction modulo p of certain two-dimensional crystalline representations*, Int. Math. Res. Not. IMRN (2009), no. 12, 2303–2317. MR 2511912 (2010g:11201)

[Buz03] Kevin Buzzard, *Analytic continuation of overconvergent eigenforms*, J. Amer. Math. Soc. **16** (2003), no. 1, 29–55 (electronic). MR 1937198 (2004c:11063)

[Buz05] ———, *Questions about slopes of modular forms*, Astérisque (2005), no. 298, 1–15, Automorphic forms. I. MR 2141701 (2005m:11082)

[CE98] Robert F. Coleman and Bas Edixhoven, *On the semi-simplicity of the $U_p$-operator on modular forms*, Math. Ann. **310** (1998), no. 1, 119–127. MR 1600034 (99b:11043)

[CM98] R. Coleman and B. Mazur, *The eigencurve*, Galois representations in arithmetic algebraic geometry (Durham, 1996), London Math. Soc. Lecture Note Ser., vol. 254, Cambridge Univ. Press, Cambridge, 1998, pp. 1–113. MR 1696469 (2000m:11039)

[Col96] Robert F. Coleman, *Classical and overconvergent modular forms*, Invent. Math. **124** (1996), no. 1-3, 215–241. MR 1369416 (97d:11090a)

[Col97] ———, *p-adic Banach spaces and families of modular forms*, Invent. Math. **127** (1997), no. 3, 417–479. MR 1431135 (98b:11047)

[Col05] ———, *The canonical subgroup of $E$ is* $\operatorname{Spec} R[x]/(x^p + \frac{p}{E_{p-1}(E,\omega)}x)$, Asian J. Math. **9** (2005), no. 2, 257–260. MR 2176608 (2006g:14074)

[Con06] Brian Conrad, *Modular curves and rigid-analytic spaces*, Pure Appl. Math. Q. **2** (2006), no. 1, part 1, 29–110. MR 2217566 (2007a:14026)

[Con07] ———, *Arithmetic moduli of generalized elliptic curves*, J. Inst. Math. Jussieu **6** (2007), no. 2, 209–278. MR 2311664 (2008e:11073)

[Del71] Pierre Deligne, *Formes modulaires et représentations $\ell$-adiques.*, Sémin. Bourbaki 1968/69, No.355, 139-172 (1971)., 1971.

[DR73] P. Deligne and M. Rapoport, *Les schémas de modules de courbes elliptiques*, Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), Springer, Berlin, 1973, pp. 143–316. Lecture Notes in Math., Vol. 349. MR 0337993 (49 #2762)

[DS05] Fred Diamond and Jerry Shurman, *A first course in modular forms*, Graduate Texts in Mathematics, vol. 228, Springer-Verlag, New York, 2005. MR 2112196 (2006f:11045)

[Dwo62] Bernard Dwork, *On the zeta function of a hypersurface*, Inst. Hautes Études Sci. Publ. Math. (1962), no. 12, 5–68. MR 0159823 (28 #3039)

[GM95] Fernando Q. Gouvêa and Barry Mazur, *Searching for p-adic eigenfunctions*, Math. Res. Lett. **2** (1995), no. 5, 515–536. MR 1359960 (96i:11048)

[Gro90] Benedict H. Gross, *A tameness criterion for Galois representations associated to modular forms (mod p)*, Duke Math. J. **61** (1990), no. 2, 445–517. MR 1074305 (91i:11060)

[HM98] Joe Harris and Ian Morrison, *Moduli of curves*, Graduate Texts in Mathematics, vol. 187, Springer-Verlag, New York, 1998. MR 1631825 (99g:14031)

[IS95] H. Iwaniec and P. Sarnak, *$L^\infty$ norms of eigenfunctions of arithmetic surfaces*, Ann. of Math. (2) **141** (1995), no. 2, 301–320. MR 1324136 (96d:11060)

[Kat73] Nicholas M. Katz, *p-adic properties of modular schemes and modular forms*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), Springer, Berlin, 1973, pp. 69–190. Lecture Notes in Mathematics, Vol. 350. MR 0447119 (56 #5434)

[Kim06] Walter Kim, *Ramification points on the eigencurve and the two variable symmetric square p-adic L-function*, ProQuest LLC, Ann Arbor, MI, 2006, Thesis (Ph.D.)–University of California, Berkeley. MR 2709141

[KM85] Nicholas M. Katz and Barry Mazur, *Arithmetic moduli of elliptic curves*, Annals of Mathematics Studies, vol. 108, Princeton University Press, Princeton, NJ, 1985. MR 772569 (86i:11024)

[Koi75] Masao Koike, *On some p-adic properties of the Eichler-Selberg trace formula*, Nagoya Math. J. **56** (1975), 45–52. MR 0382170 (52 #3058)

[Koi76] M. Koike, *On p-adic properties of the Eichler-Selberg trace formula. II*, Nagoya Math. J. **64** (1976), 87–96. MR 0429747 (55 #2757)

[Leh49] Joseph Lehner, *Further congruence properties of the Fourier coefficients of the modular invariant $j(\tau)$*, Amer. J. Math. **71** (1949), 373–386. MR 0027802 (10,357b)

[Loe07] David Loeffler, *Spectral expansions of overconvergent modular functions*, Int. Math. Res. Not. IMRN (2007), no. 16, Art. ID rnm050, 17. MR 2353090 (2009e:11084)

[Ram16] S Ramanujan, *On certain arithmetical functions*, Trans. Cambridge Philos. Soc. **22** (1916), 159–184.

[Ram06] Nick Ramsey, *Geometric and p-adic modular forms of half-integral weight*, Ann. Inst. Fourier (Grenoble) **56** (2006), no. 3, 599–624. MR 2244225 (2007c:11057)

[Ram08] ———, *The half-integral weight eigencurve*, Algebra Number Theory **2** (2008), no. 7, 755–808, With an appendix by Brian Conrad. MR 2460694 (2010a:11077)

[Ram09] ———, *The overconvergent Shimura lifting*, Int. Math. Res. Not. IMRN (2009), no. 2, 193–220. MR 2482114 (2010i:11060)

[Sar95] Peter Sarnak, *Arithmetic quantum chaos*, The Schur lectures (1992) (Tel Aviv), Israel Math. Conf. Proc., vol. 8, Bar-Ilan Univ., Ramat Gan, 1995, pp. 183–236. MR 1321639 (96d:11059)

[Ser62] Jean-Pierre Serre, *Endomorphismes complètement continus des espaces de Banach p-adiques*, Inst. Hautes Études Sci. Publ. Math. (1962), no. 12, 69–85. MR 0144186 (26 #1733)

[Ser73a] ———, *Congruences et formes modulaires [d'après H. P. F. Swinnerton-Dyer]*, Séminaire Bourbaki, 24e année (1971/1972), Exp. No. 416, Springer, Berlin, 1973, pp. 319–338. Lecture Notes in Math., Vol. 317. MR 0466020 (57 #5904a)

[Ser73b] ———, *Formes modulaires et fonctions zêta p-adiques*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, 1972), Springer, Berlin, 1973, pp. 191–268. Lecture Notes in Math., Vol. 350. MR 0404145 (53 #7949a)

[Sil86] Joseph H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics, vol. 106, Springer-Verlag, New York, 1986. MR 817210 (87g:11070)

[Sil94] ———, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics, vol. 151, Springer-Verlag, New York, 1994. MR 1312368 (96b:11074)

[SS89] A. Seeger and C. D. Sogge, *Bounds for eigenfunctions of differential operators*, Indiana Univ. Math. J. **38** (1989), no. 3, 669–682. MR 1017329 (91f:58097)

[TO70] John Tate and Frans Oort, *Group schemes of prime order*, Ann. Sci. École Norm. Sup. (4) **3** (1970), 1–21. MR 0265368 (42 #278)

[Wan98] Daqing Wan, *Dimension variation of classical and p-adic modular forms*, Invent. Math. **133** (1998), no. 2, 449–463. MR 1632794 (99d:11039)

[Wat38] G. N. Watson, *Ramanujans vermutung über zerfällungsanzahlen*, J. Reine Angew. Math. **179** (1938), 97–128.