Overconvergent Modular Forms

David Marcil

March 30th, 2021

0 - Motivation From Classical Modular Forms

When discussing modular forms over arbitrary rings, it makes more sense to use their interpretation as q-expansions or as functions on elliptic curves, then as function on the upper half-plane. This leads us to three main equivalent definitions of modular curves over a ring R and we will try to find analogous definitions in a p-adic setting afterward. We will almost only care about modular curves of level 1.

Definition (Version 1)**.** A (meromorphic) modular form f of weight k (and level 1) over R is a function on isomorphism classes (E, ω), where

(1.1) $E \to \mathrm{Spec}(A)$ is an elliptic curves over some R-algebra A.

(1.2) $\omega \in H^0(E, \Omega^1_{E/A})$ is a nowhere vanishing differential

(2.1) $f(E/A, \omega) \in A$

(2.2) $f(E/A, \lambda\omega) = \lambda^{-k} f(E/A, \omega)$, for all $\lambda \in A^\times$.

(2.3) f commutes with pullbacks

Note that the hypothesis (1.1) and (1.2) are about the domain of f and the other 3 are about the properties of f.

Definition (Version 2)**.** A (merormophic) modular form f of weight k over R is a global section

$$f \in H^0(Y(1)_R, \omega_H^{\otimes k}) \ ,$$

where ω_H is the base changed to R of the Hodge line bundle over $Y(1)$. We just drop the "meromorphic" adjective if f extends to a global section of $H^0(X(1)_R, \omega_H^{\otimes k})$.

Definition (Version 3)**.** Although it is uncommon to define (meromorphic) modular formss as particular q-expansions, it is worth noting that a modular curve f as above over R is uniquely determined by its q-expansion

$$f(Tate(q), \omega_{can}) \in R((q)) \ ,$$

where $Tate(q)$ is the Tate curve over $\mathbb{Z}((q))$ (base changed to R) with its canonical non-vanishing differential ω_{can}. We have that f is actually a modular form if and only if $f(Tate(q), \omega_{can}) \in R[[q]]$.

For more details on this subject, see Haodong's notes from his talks on p-adic properties of modular schemes and modular forms.

1 - Serre's Approach to p-adic Modular Forms

Fix a prime integer $p \geq 2$. One way to define p-adic modular forms is as limits of classical modular forms, where the coefficients converge *uniformely* p-adically. Over $\mathbb{Z}_p$, one can juste define a q-expansion

$$f = \sum_{n \geq 0} a_n q^n \in \mathbb{Q}_p[[q]]$$

to be a p-adic modular forms if there exists a sequence (f_i), such that f_i is a modular form of weight k_i (and level 1) with rational q-expansion and

$$||f - f_i||_p := \sup_n |a_n(f) - a_n(f_i)|_p \xrightarrow{i \to +\infty} 0$$

Note that this implies that the p-valuations of the coefficients of f are bounded below, hence we in fact have $f \in \mathbb{Z}_p[[q]] \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$.

Theorem 1. *Let f, f_i and k_i be as above. The sequence $(k_i)_i$ converges in to some element k in*

$$\varprojlim_m \mathbb{Z}/(p^m(p-1))\mathbb{Z} = \mathbb{Z}_p \times \mathbb{Z}/(p-1)\mathbb{Z} \ (\cong \mathbb{Z}_p^\times)$$

Moreover, if another sequence of modular forms $(g_i)_i$ converges to f in the sense above, the limit of the weights of g_i is k again. We say that f has weight $k \in \mathbb{Z}_p \times \mathbb{Z}/(p-1)\mathbb{Z}$.

One of the most well-known example of a non-classical p-adic modular form comes from congruencees of Eisenstein series.

Example 1. Let $k \geq 4$ be any integer. The classical Eisenstein series of weight k is

$$G_k = \frac{-B_k}{2k} + \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n \ ,$$

where $\sigma_r(n) = \sum_{d|n} d^r$ and B_k is the k-th Bernouilli number.

Now, take a sequence of integers (k_i), such that all $k_i \geq 4$, $k_i \to +\infty$ and the sequences converges to some $k \in \mathbb{Z}_p \times \mathbb{Z}/(p-1)\mathbb{Z}$. Assume for simplicity that $p-1$ does not divide k_i for i large enough or equivalently, $k \notin \mathbb{Z}_p \times \{0\}$.

Using Fermat's Little Theorem, it is easy to see that

$$\sigma_{k_i - 1}(n) \to \sigma_{k-1}(n)^* := \sum_{p \nmid d | n} d^{k-1} \in \mathbb{Q}_p$$

where this sequence converges with respect to the natural topology of $\mathbb{Q}_p$. Note that the divisors $d \mid n$ that are not coprime to p will contribute to 0 in the limit and the other divisors $p \nmid d \mid n$ are units in $\mathbb{Z}_p$, so the exponent d^k makes sense.

The all-important *Kummer congruences* tells us that the constant term $a_0(G_{k_i})$ also converges to some $a_0^* \in \mathbb{Q}_p$. It follows that we have a *non-classical* p-adic modular form

$$G_k^* := a_0^* + \sum_{n \geq 1} \sigma_{k-1}^*(n) q^n$$

A theorem of Serre tells us that one does not actually need to prove the Kummer congruences (which is a difficult task) to obtain this p-adic modular forms. Indeed, one can use the following result to extend congruences of higher order terms to the 0-th term of the q-expansion

Theorem 2. *Let $f^{(i)} = \sum_{n \geq 0} a_n^{(i)} q^n$ be a p-adic modular forms of weight k_i such that*

1. *The sequence $(k_i)_i$ converges to some k in $\mathbb{Z}_p \times \mathbb{Z}/(p-1)\mathbb{Z}$.*

2. *The sequence $(a_n^{(i)})_i$ converges to some $a_n \in \mathbb{Q}_p$, uniformely in n.*

In that case, the sequence $(a_0^{(i)})_i$ also converges to some $a_0 \in \mathbb{Q}_p$ and $f = \sum_{n \geq 0} a_n q^n$ is a p-adic modular forms of weight k.

For more details on this example, see Vonk's notes [Von19] or even better, Serre's original article [Ser73].

Remark 1. The example above, together with Serre's theorem, already show how useful p-adic modular forms can be when studying congruences of classical modular forms. Furthermore, the coefficent $a_0(G_k^*) = a_0^*$ as a function of $k \in \mathbb{Z}_p \times \mathbb{Z}/(p-1)\mathbb{Z}$ is actually intimately related to the Kubota-Leopoldt p-adic L-function. This gives a surprising way to construct it.

2 - Katz's Approach to p-adic Modular Forms

Although Serre's approach is already quite powerful and impressive, it is not really flexible enough to generalize to other rings. As we mentioned in section 0, q-expansions might not be the most natural way to build the theory of (p-adic) modular forms. We now approach the theory using (simultaneously) the other two definitions of classical modular forms over a ring R. Before accomplishing this, we need to discuss the Hasse Invariant and some more geometry.

2.1 - Hasse Invariant and More General p-adic Modular Forms

The Hasse invariant allows us to detect, in a general sense, elliptic curves whose reduction are supersingular. We will later use its lift to explain how some elliptic curves are "more supersingular" than others. Let S be a ring such that $p = 0$ in S.

Definition. The *Hasse invariant* A is a (classical) modular form over S of weight $p-1$. It's exact construction will not matter to us, for more details see [Cal13], [Kat73] or [Von19]. Note that the Hasse invariant A depends on S, even though the notation does not show this dependence.

The important properties for us are the following :

1. If $S = \mathbb{F}_p$ is a field, then $A(E, \omega)$ is 0 if and only if E is supersingular.

2. The q-expansion of A is $A(Tate(q), \omega_{can}) = 1$.

3. Suppose there exists some ring R of characteristic 0 such that $S = R/pR$. If $p \geq 5$, the Eisenstein series E_{p-1} over R reduces to A in S. Otherwise, if $p = 2$, then $E_4 \equiv A^4 \bmod 8$ and, finally, if $p = 3$, then $E_6 \equiv A^3 \bmod 9$. Here, $E_k = \frac{-2k}{B_k} G_k$ is the Eisenstein series, normalized to have constant term 1.

Let R be a p-adically completed ring.

Definition. A p-adic modular form f of weight k (and level 1) is a function on isomorphism classes of pairs (E, ω) such that

(1.1) $E \to \mathrm{Spec}(A)$ is an elliptic curve, where A is a p-adically completed R-algebra.

(1.2) $\omega \in H^0(E, \Omega^1_{E/A})$ is a non-vanishing differential on E.

(1.3) The Hasse invariant $A(E/B, \omega_B)$ is invertible, $B = A/pA$.

(2.1) $f(E/A, \omega) \in A$

(2.2) $f(E, \lambda\omega) = \lambda^{-k} f(E, \omega)$, for all $\lambda \in A^\times$.

(2.3) f commutes with pullback

(2.4) $f(Tate(q), \omega_{can}) \in A[[q]]$.

The space of such function is denoted by $M_k(R, 0)$ $(= M_k(\Gamma(1), R, 0))$. The inclusion of the "0" in the notation will make sense shortly, after we introduce *overconvergent* modular forms. Note that once more, the first 3 conditions are about the domain of f and the last 4, about its properties.

We pick up the third condition for its domain because, in the case $R = \mathbb{Z}_p$, we recover Serre's p-adic modular curves and interpreted as functions on elliptic curves, they might not be well-defined on supersingular elliptic curves. This observation will exactly be what leads us to consider *overconvergent* modular forms.

Remark 2. Geometrically, p-adic modular forms are global sections of powers of the Hodge line bundle on the ordinary locus $X^{rig}[0]$ as drawn below (where $X = X(1)$) :

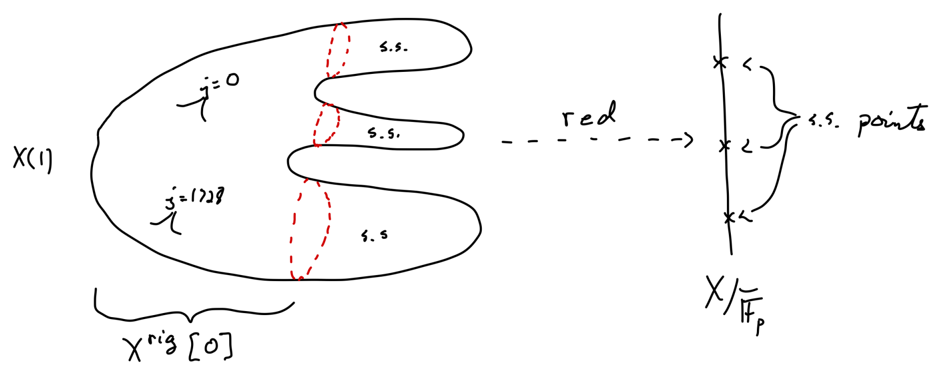

Figure 1: Rigid analytic ordinary locus

Here, we use the reduction map red : $X(\mathbb{C}_p) \to X(\bar{\mathbb{F}}_p)$. This maps exists since X is proper, hence all $\mathbb{C}_p$-points extend uniquely to $\mathcal{O}_{\mathbb{C}_p}$-points, which can be reduced to $\bar{\mathbb{F}}_p$-points. The picture suggest that $X^{rig}[0]$ should be defined as the complement of the pre-image of the supersingular points of $X(\bar{\mathbb{F}}_p)$ along the reduction map.

2.2 - Geometry Behind p-adic Modular Forms - Rigid Analytic Spaces

We now explain more about the *rigid analytic space* $X^{rig}[0]$ drawn above. The drawing suggest that we have "(open) disks" above all the supersingular points of $X(1)/\mathbb{F}_p$ and we want to remove them from $X(1)$. Unfortunately, open disks are not really objects in the theory of schemes and hence we certainly can't remove them. This is where rigid analytic geometry comes into play. We won't explain the details of the theory but introduce the relevant facts.

Let K be a complete non-Archimedian field with (non-trivial) absolute value $|\cdot|$, e.g. $K = \mathbb{Q}_p$ or $\mathbb{C}_p$.

Definition. A *Tate algebra* over K is

$$K\langle T_1, \ldots, T_n\rangle := \left\{ \sum_I a_I X^I : a_I \in K, |a_I| \to 0 \text{ as } |I| \to +\infty \right\}$$

where $I = (i_1, \ldots, i_n)$, $i_j \geq 0$ and $|I| = \sum i_j$.

This is the equivalent of affine space in rigid analytic geometry. We associate to it the space

$$\text{Specmax}(K\langle T_1, \ldots, T_n \rangle) = \bar{R}^n$$

where $\bar{R}$ is the ring of integers of the algebraic closure $\bar{K}$ of K. Namely, we associate an n-dimensional ball of radius 1 to this Tate algebra.

Definition. A *K-affinoid algebra* is a quotient of a Tate algebra over K and an *affinoid* is the corresponding set of maximal ideals.

One can equip such an affinoid with a *Grothendieck topology*, called the *G-topology*, and rigid analytic space is one formed by gluing such affinoids, together with its own structure sheaf with respect to the G-topology. It is important to keep in mind that its is not an actual topology on the space (such as the Zariski topology) but more of a *site* (such as the étale site). It allows one to have a more general theory of sheaf in this setting and is in some sense finer than the Zariski topology. Most importantly, it allows one to substract (open) balls (of various radius), as the ones above.

Theorem 3. *Given a variety X over K, we can associate to it a rigid analytic space X^{rig} such that the geometric space associated to X^{rig} is the set of closed points of X*

We will mostly be interested in the case $X = X(1)$, over $K = \mathbb{Q}_p$ or $\mathbb{C}_p$, and the ordinary locus

$$X^{rig}[0] = X^{rig} \setminus \{\text{rigid analytic open disks above supersingular points}\}$$

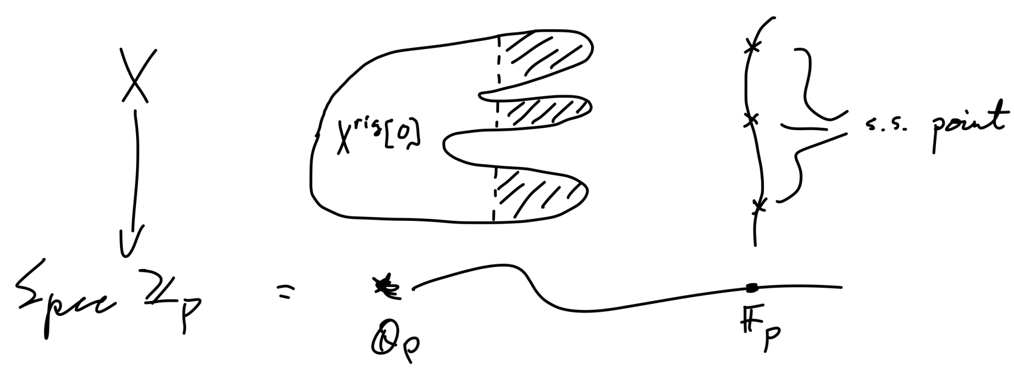

Figure 2: Rigid analytic ordinary locus

Now that this is introduced, we can rephrase our definitions of p-adic modular forms over K in a slightly more concice manner

Definition. A p-adic modular form f of weight k (and level 1) is a global section

$$f \in H^0(X(1)_K^{rig}[0], \omega_H^{\otimes K})$$

Example 2. If $p = 2$, then $j_E = 0$ is the only super singular elliptic j-invariant. In other words, the ordinary locus can be described as $|j|_2 \geq 1$ or equivalently, $|j^{-1}|_2 \leq 1$. From that, one can argue that p-adic modular forms over $\mathbb{C}_2$ of weight 0 are all functions in the variable j^{-1} that converge on $|j^{-1}|_2$.

In other words, $M_0(\mathbb{C}_2, 0)$ is the Tate algebra $\mathbb{C}_2\langle j^{-1} \rangle$ and the 2-adic ordinary locus $X^{rig}[0]$ is the corresponding ball.

Remark 3. Although this space is very interesting, it is too big in some sense. One way to see this is through the spectral theory of the Hecke operator U_p on $M_k(R, 0)$. Namely, one can check that all values $\lambda \in R$ with positive valuation are eigenvalues of U_p (see [Cal13] for details).

Therefore, the spectrum of U_p is continuous and quite large, so there is no hope of finding an Hilbert basis of that space formed of Hecke (generalised) eigenform, as in the classical case. We now try to restrict ourselves to somewhat smaller spaces (that are still infinite dimensional) where we the spectral theory of U_p is a little more interesting.

2.3 - Overconvergent Modular Forms

To introduce *overconvergent* p-adic modular forms, we first need to construct rigid analytic spaces that (strictly) contains $X^{rig}[0]$. For that, we need to generalize our definition of Hasse invariant.

Let $x \in X^{rig}(\mathbb{C}_p)$ and let (E, ω) be the elliptic curve (over $\mathbb{C}_p$) it corresponds to. We can find a lift $\tilde{A}$ to $\mathbb{C}_p$ of the Hasse invariant A (over $\mathbb{F}_p$) in a small neighborhood of x. If $p \geq 5$, then E_{p-1} is in fact a global lift. The valuation $v_p(\tilde{A}(E, \omega))$ is independant of our choice of lift $\tilde{A}$, provided it is less or equal to 1.

In other words, the quantity

$$\min\{v_p(\tilde{A}(E, \omega), 1\}$$

is an invariant of x, called its *Hasse invariant.* We now have a map

$$X^{rig}(\mathbb{C}_p) \to [0, 1]$$

defined as evaluation of the Hasse invariant and, given $r \in [0, 1]$, we set $X^{rig}[r]$ as the rigid analytic space coming from the pre-image of $[0, r]$. This is obviously consistent with our previous notation if $r = 0$.

Remark 4. This amounts to taking (open) disks of smaller radius over the supersingular points of $X(\bar{F}_p)$. We can do this because the full pre-image is a disk of radius 1 associated to a Tate algebra in 1 variable and we can make the radius smaller by taking the affinoid associated to the Tate algebra where the local parameter is rescaled.

Definition. An *overconvergent p-adic modular form* f of weight k and radius $r \in [0, 1]$ is

$$f \in H^0(X^{rig}[r], \omega^{\otimes k})$$

and this space is denoted $M_k^\dagger(r)$ $(= M_k^\dagger(\Gamma(1), r))$.

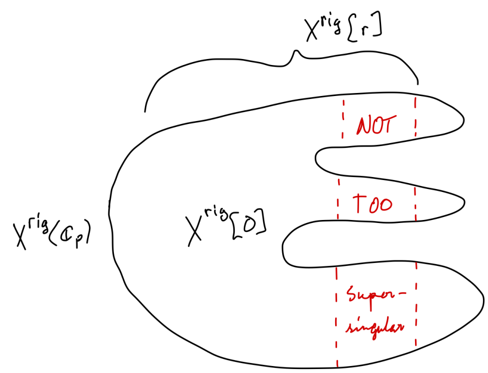

Figure 3: Drawing of $X^{rig}[r]$ for $0 < r < 1$

Example 3. We mentioned earlier that

$$M_0(\Gamma(1), 0) = M_0(\Gamma(1), \mathbb{C}_2) = \mathbb{C}_2\langle j^{-1}\rangle$$

We now want the part that "overconverges" beyond the ordinary locus, up to an annulus of radius r over the unique supersingular $j = 0$. This amounts to forcing a certain speed of decay of the coefficients of the elements of $\mathbb{C}_2\langle j^{-1}\rangle$.

Note that $j = \frac{E_4^3}{\Delta}$, where Δ is the usual discriminant modular form of weight 12, and as we saw previously, E_4 is a lift of A^3. On the supersingular disk, elliptic curves obviously have good reduction, i.e. $v(\Delta) = 0$. From this, we obtain that

$$v(j) = v(E_4^3) = 3v(E_4) = 12v(A) \ ,$$

and hence $v(A) \leq r \iff v(j) \leq 12r \iff v(j^{-1}) \geq -12r$.

Therefore, we want power series in j^{-1} that converge on $|j^{-1}| \leq 2^{12r}$, i.e.

$$M_0^\dagger(r) = \{\sum_{n\geq 0} a_n j^{-n} \colon |a_n|_2(2^{12r})^n \to 0\} = \mathbb{C}_2\langle 2^{12r}j^{-1}\rangle$$

Remark 5. There is a way to describe the above even more nicely when working on $X_0(2)$ instead of $X(1)$. Namely,

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

where $f = \Delta(2z)/\Delta(z)$ is the Hauptmodul of $X_0(2)$. This is related to the existence of a canonical subgroup of order 2 for elliptic curves that are not "too" supersingular over $\bar{\mathbb{F}}_2$ giving us a "partial" section to the natural map $X_0(2) \to X(1)$.

Theorem 4. *There is a canonical inclusion $M_k(\Gamma_0(p^n)) \subset M_k^\dagger(r)$, for $r > 0$.*

Remark 6. Note that this is not obvious since modular forms in $M_k(\Gamma_0(p))$ need to be evaluated on elliptic curves equipped with some level structure. If E is not "too" supersingular, i.e. r is small enough, there is a canonical choice of level structure on E and we use it in our evaluation of the desired overconvergent p-adic modular form.

2.4 - Notes on Spectral Theory of U_p

Now that we have introduced the space of overconvergent modular forms, we ask : Is it really "better" than the space Serre introduced? To answer this, we give quick note about the spectral theory of U_p.

Theorem 5. *The operator U_p extends to*

$$M_k^\dagger(r) \to M_k^\dagger(pr) \xrightarrow{can} M_k^\dagger(r)$$

provided $r < \frac{p}{p+1}$. Moreover, it is compact.

Once more, this has to do with the canonical level p structure that can be given to elliptic curves that are not "too" supersingular.

Conjecture 1 (Spectral Conjecture)**.** *Suppose that $r \in (\frac{1}{p+1}, \frac{p}{p+1})$. Then, any $F \in M_k^\dagger(r)$ has a spectral decomposition, i.e.*

$$F = \sum_i c_{i,F} v_i$$

where the v_i's are generalized eigenform U_p and the infinite sum converges with respect to some (strong) norm on $M_k^\dagger(r)$.

Theorem 6 (Loeffler, 07). *The Spectral conjecture is ture if* $p = 2$ *and* $r \in (5/12, 7/12)$. *Moreover, the (strong) norm of* $F \in M_k^\dagger(r)$ *is satisfies some further properties.*

Remark 7. The v_i's above are *generalized* eigenform. The question of whether or not they are actually eigenforms, i.e. U_p is semisimple, is an interesting topic in this theory.

Definition. Suppose $F \in M_k^\dagger(r)$, for $r \leq \frac{p}{p+1}$, is an eigenform of U_p. We define the *slope* of F has the p-valuation of its U_p-eigenvalue.

Theorem 7. *If* $F \in M_k(\Gamma_0(p))$ *is an* U_p*-eigenform, it has slope at most* $k-1$. *If* $F \in M_k^\dagger(r)$ *has slope less or equal to* $k-2$, *then it lies in* $M_k(\Gamma_0(p))$. *If it has slope* $k-1$, *then it need not be classical.*

3 - Some Computations

Consider the case $p = 2$. Obviously, j^{-1} is overconvergent, for all $r \neq 1$. It therefore makes sense to compute $U_2 j^{-1}$, at least for $r < 2/3$.

One can check that

$$U_2 j^{-1} = -744 j^{-1} - 140914688 j^{-2} - 16324041375744 j^{-3} + \dots$$

and more importantly, the 2-valuations of the coefficients are

$$3, 12, 20, 28, 35, 46, 52, 60, 67, 76, \dots$$

and at a very large scale, those look like values on the line of slope 8 (passing through the origin). Note that $8 = 12 \cdot (\frac{2}{3})$, so the coefficients grow as fast as they could while still convergins on the space $X^{rig}[\frac{2}{3}]$.

Remark 8. In terms of working with f, the Hauptmodul of $X_0(2)$, instead of j^{-1}, we would instead compute

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

and this again relates to the existence of a canonical section $X_0(2)^{rig}[\frac{2}{3}] \to X^{rig}[\frac{2}{3}]$.

References

[Cal13] F. Calegari. Congruences between modular forms. *In Arizona Winter School*, pages 1–37, 2013.

[Kat73] N. Katz. *p-adic properties of modular schemes and modular forms*, volume 350 of LNM of *In P. Deligne and W. Kuyk, editors, Modular forms in one variable III*, pages 69–190. Springer-Verlag, 1973.

[Ser73] J.-P. Serre. *Formes modulaires et fonctions zeta p-adiques*, volume 350 of LNM of *In P. Deligne and W. Kuyk, editors, Modular forms in one variable II*, pages 191–268. Springer-Verlag, 1973.

[Von19] J. Vonk. Overconvergent modular forms and their explicit arithmetic. *In Bordeaux Conference on Iwasawa Theory*, pages 1–20, 2019.