Chapter I

p-adic Modular Forms

We begin by giving an overview of the basic theory of p-adic modular forms, following, for the most part, the approach of Katz. First, we explain Katz's "p-adic modular forms with growth conditions" and how they relate to Serre's version of the theory. Then, we go on to discuss "generalized p-adic modular functions" (which, as the name suggests, include all the objects defined previously). The main references for this chapter are the foundational papers of Serre and Katz, especially [Se73], [Ka73], [Ka76], and [Ka75b].

To understand the definition of p-adic modular forms as functions of elliptic curves with extra structure (which is what we mean by a "modular" definition), we should first recall how to interpret classical modular forms in these terms.

Classical (meromorphic) modular forms can be interpreted as functions of triples $(\mathrm{E}/A, \omega, \imath)$, composed of an elliptic curve E over A, a non-vanishing invariant differential ω on E and a level structure $\imath$, obeying certain transformation laws. Equivalently, they can be thought of as global sections of certain invertible sheaves over the moduli space of elliptic curves with the given kind of level structure, which is of course a modular curve. If we restrict A to be an algebra over the complex numbers, then it is easy to see that this is equivalent to the classical theory, since the quotient of the upper half-plane by a congruence subgroup classifies elliptic curves with a level structure. The fact that allowing A to run over R-algebras produces the classical theory of "modular forms defined over R" (i.e., complex modular forms whose Fourier expansion coefficients belong to R) is known as the "q-expansion principle". An exposition of the classical theory in this spirit can be found in the first chapter of [Ka73]; the first appendix to the same paper explains how this formulation relates to the classical definitions.

We wish to obtain a p-adic theory of modular forms, i.e., a theory which reflects the p-adic topology in an essential way. This cannot be done by simply mimicking the definition of classical modular forms as functions of elliptic curves with level structure and differential, because the space of "classical modular forms over $\mathbf{Z}_p$" thus obtained is simply the tensor product with $\mathbf{Z}_p$ of the space of classical modular forms over $\mathbf{Z}$. To obtain a properly p-adic theory, we should take into account the p-adic topology by allowing limits of classical forms. This can either be done directly in terms of Fourier expansions of modular forms, or one can try to obtain a "modular" definition.

The first approach, which is due to Serre (in [Se73]), is to *identify* a classical modular form with the set of its q-expansions, and then to consider limits by using the p-adic topology on $\mathbf{Z}_p[[q]]$. One is then able to show that whenever a sequence of classical forms tends to a limit, their weights tend to a "p-adic weight" χ, which is just a character $\chi : \mathbf{Z}_p^\times \longrightarrow \mathbf{Z}_p^\times$. This produces an elementary theory with strong ties to the theory of congruences between classical modular forms, which turns out to be a special case of the "modular" theory developed by Katz in [Ka73].

To obtain a "modular" theory of p-adic modular forms, one should define them as functions on elliptic curves, or, equivalently, as sections of bundles on a modular curve. This is achieved by Katz's idea of considering the rigid analytic space obtained by deleting p-adic disks around the supersingular points in the (compactified) moduli space $\mathcal{M}(\mathrm{N})$ of elliptic curves with a $\Gamma_1(\mathrm{N})$-structure over $\mathbf{Z}_p$. To do this, we recall that, for $p \geq 5$, the classical modular form E_{p-1} is a p-adic lifting of the Hasse invariant (this is equivalent to the well-known fact that the q-expansion of E_{p-1} is congruent to 1 modulo p, see [Ka73, Section 2.1]), and consider regions on $\mathcal{M}(\mathrm{N})$ where E_{p-1} is "not too near zero". Since we would like to remove as small a disk as possible, we will allow the meaning of "not too near zero" to vary in terms of a parameter r, which we call the "growth condition". Taking $r = 1$ amounts to restricting ourselves to ordinary curves (i.e., to deleting the supersingular disks completely), and the resulting theory is the same as Serre's. However, if r is not a p-adic unit, we get a smaller space of "overconvergent" forms, which can be evaluated at curves which are "not too supersingular". Many of the interesting questions of the theory turn on the relation between these spaces as one varies r. The idea of considering modular forms with growth conditions at the supersingular points seems to be originally due to Dwork (for example, in [Dw73]); it was first developed systematically by Katz in [Ka73].

The idea of deleting the supersingular curves (more precisely, the curves with supersingular reduction) may sound strange at first, but is in fact quite natural in the context of what is wanted. We would like congruences of q-expansions to reflect congruences of the modular forms themselves, so that, for example,

$$\mathrm{E}_{p-1}(q) \equiv 1 \pmod{p}$$

in $\mathbf{Z}_p[[q]]$ should imply that there exists a modular form f such that

$$\mathrm{E}_{p-1} - 1 = pf;$$

this, however, is certainly false if we allow f to be evaluated at (a lifting of) a supersingular curve since the value of E_{p-1} at any such curve must be divisible by p (because it lifts the Hasse invariant). It turns out that omitting the supersingular disks (and possibly restricting the choice of differential on the curve) does the trick, and congruence properties of q-expansions are then reflected in congruence properties of p-adic modular forms. It also turns out that it is interesting to vary the radius of the omitted disk, introducing growth conditions and making the theory richer.

The choice of E_{p-1} as a lifting of the Hasse invariant automatically restricts us to primes $p \geq 5$. For $p = 2$ and $p = 3$, one must choose a different lifting of the Hasse invariant (of higher level, since there is no lifting of level 1); one knows that such liftings exist for $p = 3$ and $N \geq 2$ and for $p = 2$ and $3 \leq N \leq 11$. This means that it is possible to construct a theory on the same lines for $p = 3$ and any level $N \geq 2$ (and then obtain a level $N = 1$ theory by taking the fixed points under the usual group action); for $p = 2$, however, one will only get a theory for levels divisible by some number between 3 and 11, and again try to use group actions to get the full theory. In any case, since we will later need to restrict our theory to the case $p \geq 7$ (for the spectral theory of the U operator), we have preferred to avoid these questions entirely by stating our results only for primes $p \geq 5$. In [Ka73], Katz discusses these problems further, and takes the cases $p = 2$ and $p = 3$ into account in the statements of his theorems.

I.1 Level Structures and Trivializations

In what follows, p will denote a fixed rational prime, $p \geq 5$, and N a fixed level with $(N, p) = 1$. To guarantee that the moduli problems under consideration are representable, we will often assume that $N \geq 3$ (especially when discussing forms with growth conditions). We let B denote a "p-adic ring", i. e., a $\mathbf{Z}_p$-algebra which is complete and separated in the p-adic topology. In most cases, B will be a p-adically complete discrete valuation ring or a quotient of such.

Let E be an elliptic curve over a p-adic ring B. We will consider, following Katz, level structures on E of the following kind:

Definition I.1.1 *Let* E *be an elliptic curve over B, and let* $E[Np^\nu]$ *denote the kernel of multiplication by* Np^ν *on* E, *considered as a group scheme over B. An arithmetic level* Np^ν *structure, or* $\Gamma_1(Np^\nu)^{arith}$*-structure, on* E *is an inclusion*

$$\imath : \mu_{Np^\nu} \hookrightarrow E[Np^\nu]$$

of finite flat group schemes over B.

It is clear that, if $\nu > 0$, the existence of such an inclusion implies that E is fiber-by-fiber ordinary, so that we are automatically restricting our theory to such curves whenever the level is divisible by p.

We will denote by $\mathcal{M}^{o}(Np^\nu)$ the moduli space of elliptic curves with an arithmetic level Np^ν structure. When $\nu > 0$, this is an open subscheme of the moduli space of elliptic curves with a $\Gamma_1(Np^\nu)$-structure as defined by Katz and Mazur in [KM85], which we denote by $\mathcal{M}_1(Np^\nu)$. This last space may be compactified by adding "cusps" (see [KM85], [DeRa]); this produces a proper scheme which we will denote by $\overline{\mathcal{M}}_1(Np^\nu)$, which contains as an (affine when $\nu > 0$) open subscheme the scheme obtained by adding the cusps to $\mathcal{M}^o(\mathrm{N}p^\nu)$, which we denote by $\mathcal{M}(\mathrm{N}p^\nu)$. Thus we have a diagram

$$\begin{array}{ccc} \mathcal{M}^o(\mathrm{N}p^\nu) & \hookrightarrow & \mathcal{M}_1(\mathrm{N}p^\nu) \\ \downarrow & & \downarrow \\ \mathcal{M}(\mathrm{N}p^\nu) & \hookrightarrow & \overline{\mathcal{M}}_1(\mathrm{N}p^\nu) \end{array}$$

Moreover, the horizontal arrows are isomorphisms outside the primes dividing $\mathrm{N}p^\nu$; in particular, they are isomorphisms of schemes over $\mathbf{Z}_p$ when $\nu = 0$. (For more comments on these moduli spaces, see [MW86].)

We will say an arithmetic level $\mathrm{N}p^\nu$structure is *compatible* with an arithmetic level $\mathrm{N}p^\mu$ structure if the obvious diagram of inclusions commutes.

We will also want to consider "trivialized elliptic curves over B", which we define as follows:

Definition I.1.2 *Let* E *be an elliptic curve over B. A trivialization of* E *is an isomorphism*

$$\varphi : \hat{\mathrm{E}} \xrightarrow{\sim} \hat{\mathbf{G}}_m$$

of formal groups over B, where $\hat{\mathrm{E}}$ *denotes the formal completion of* E *along its zero-section. A trivialized elliptic curve over B is an elliptic curve* E *over B together with a trivialization* φ.

It is clear that such an isomorphism can only exist when E is fiber-by-fiber ordinary; conversely, given such an E, one can obtain a trivialization after a base change (see [Ka75b]).

We will say that a $\Gamma_1(\mathrm{N}p^\nu)^{arith}$-structure $\imath : \mu_{\mathrm{N}p^\nu} \hookrightarrow \mathrm{E}[\mathrm{N}p^\nu]$ is compatible with a trivialization φ if the induced map

$$\mu_{p^\nu} \hookrightarrow \hat{\mathrm{E}} \xrightarrow{\sim} \hat{\mathbf{G}}_m$$

is the canonical inclusion.

It is important to notice that a trivialization φ determines a sequence of mutually compatible $\Gamma_1(p^\nu)^{arith}$-structures. In fact, since there is an equivalence between the categories of p-divisible smooth connected commutative formal groups over B and of connected p-divisible groups over B (see [Ta67]), giving a trivialization is *equivalent* to giving such a sequence of level structures (which we might call a $\Gamma_1(p^\infty)^{arith}$-structure), and we will use either without further comment.

I.2 p-adic Modular Forms with Growth Conditions

In this section we define and review the basic properties of Katz's p-adic modular forms with growth conditions, which include as a special case Serre's p-adic modular forms. All of this is due to Katz in [Ka73], to which we will constantly refer for more details.

I.2.1 Definitions

Let B be a p-adic ring and let $r \in B$; we will usually assume B is either a p-adically complete discrete valuation ring or a quotient of such a ring. A p-adic modular form with growth condition r will be a function on "test objects with growth condition r" with prescribed transformation laws, by analogy with classical modular forms considered as functions of test objects consisting of elliptic curves with a level structure and a non-vanishing differential. Our test objects will be elliptic curves with level structures and a non-vanishing differential plus an extra structure which guarantees that a p-adic lifting of the Hasse invariant of the curve is "not too near zero".

Definition I.2.1 *Let A be a p-adically complete and separated B-algebra. A test object of level* N *and growth condition r defined over B is a quadruple $(\mathrm{E}/A, \omega, \imath, Y)$, where* E *is an elliptic curve over A, ω is a nonvanishing differential on* E*, $\imath$ is a $\Gamma_1(\mathrm{N})^{arith}$-structure on* E*, and $Y \in A$ satisfies*

$$Y \cdot \mathrm{E}_{p-1}(\mathrm{E}, \omega) = r.$$

We will sometimes refer to Y as an "r-structure" on the elliptic curve E. It clearly can only exist when the p-adic valuation of $\mathrm{E}_{p-1}(\mathrm{E}, \omega)$ is smaller than the p-adic valuation of r; for example, if $r = 1$, the existence of Y implies that $\mathrm{E}_{p-1}(\mathrm{E}, \omega)$ is a unit in A. Furthermore, if B is flat over $\mathbf{Z}_p$ and we fix a differential ω on E, then the r-structure Y is uniquely defined when it exists. Thus, requiring an r-structure restricts us to curves which are "not too supersingular" (when $r = 1$, to ordinary curves).

Let $k \in \mathbf{Z}$ be an integer; then we define:

Definition I.2.2 [Katz] *A p-adic modular form of weight k, level* N*, and growth condition r defined over B is a rule f which assigns to a test object $(\mathrm{E}/A, \omega, \imath, Y)$ of level* N *and growth condition r defined over B an element*

$$f(\mathrm{E}/A, \omega, \imath, Y) \in A$$

satisfying the following conditions:

i. $f(\mathrm{E}/A, \omega, \imath, Y)$ *depends only on the isomorphism class of the triple* $(\mathrm{E}/A, \omega, \imath, Y)$,

ii. the formation of $f(\mathrm{E}/A, \omega, \imath, Y)$ *commutes with base change,*

iii. for any $\lambda \in A^\times$ *we have*

$$f(\mathrm{E}/A, \lambda\omega, \imath, \lambda^{1-p}Y) = \lambda^{-k} f(\mathrm{E}/A, \omega, \imath, Y).$$

We denote by $\mathsf{F}(B, k, \mathrm{N}; r)$ the space of all p-adic modular forms of weight k, level N, and growth condition r defined over B. The restriction to p-adic rings A in the definitions implies that we have

$$\varprojlim_n \mathsf{F}(B/p^n B, k, \mathrm{N}; r) = \mathsf{F}(B, k, \mathrm{N}; r).$$

We say a p-adic modular form with growth condition r is *overconvergent* if r is not a unit in B.

Let $f \in \mathsf{F}(B,k,\mathrm{N};r)$. Let Tate$(q)$ be the Tate elliptic curve defined over $\widehat{B((q))}$ (the p-adic completion of the ring of Laurent series with coefficients in B), let $\imath_{can}$ denote its canonical arithmetic level N structure and let ω_{can} denote its canonical differential. (For the definitions, see [Ka73, Appendix 1] or [DeRa, Chap. VII].) Since Tate(q) is ordinary, $\mathrm{E}_{p-1}(\mathrm{Tate}(q),\omega_{can})$ is invertible, so that we can take

$$Y = r \cdot \mathrm{E}_{p-1}(\mathrm{Tate}(q),\omega_{can})^{-1}$$

and evaluate f on the test object $(\mathrm{Tate}(q),\omega_{can},\imath_{can},Y)$ to obtain an element of $\widehat{B((q))}$ (the p-adic completion of $B((q))$). We call

$$f(q) = f(\mathrm{Tate}(q),\omega_{can},\imath_{can},Y)$$

the q-expansion of f, and often, as here, denote it by $f(q)$. We say that f is holomorphic if

$$f(\mathrm{Tate}(q),\omega_{can},\imath,Y) \in B[[q]]$$

for every arithmetic level N structure $\imath$ on Tate(q), and we denote the space of all such f by $\mathsf{M}(B,k,\mathrm{N};r)$. (The reader will note that our notation differs from that of Katz in [Ka73].) In this case we again have

$$\varprojlim_n \mathsf{M}(B/p^nB,k,\mathrm{N};r) = \mathsf{M}(B,k,\mathrm{N};r).$$

One may also define the subspace of cusp forms, analogously to the classical case, by requiring that the q-expansions at all the cusps belong to $qB[[q]]$, i.e., we say that $f \in \mathsf{M}(B,k,\mathrm{N};r)$ is a cusp form if

$$f(\mathrm{Tate}(q),\omega_{can},\imath,Y) \in qB[[q]]$$

for any level N structure $\imath$ on the Tate curve. We then denote the space of all such by $\mathsf{S}(B,k,\mathrm{N};r)$.

Remark: Alternatively, we can think of modular forms as global sections of the invertible sheaves $\underline{\omega}^{\otimes k}$, where $\underline{\omega}$ denotes the sheaf on $\mathcal{M}(\mathrm{N})$ obtained by pulling back the sheaf of differentials on the universal elliptic curve on $\mathcal{M}(\mathrm{N})$ via the zero-section. Then we should take as test objects triples $(\mathrm{E}/S,\imath,Y)$, where E is an elliptic curve over a scheme S, $\imath$ is an arithmetic level N structure, and Y is a global section of $\underline{\omega}^{\otimes(1-p)}$ satisfying $Y\cdot \mathrm{E}_{p-1}(\mathrm{E}/S,\imath) = r$ (where of course we view E_{p-1} as a global section of $\underline{\omega}^{\otimes(p-1)}$). This accounts for the need to change both Y and ω in the transformation rule above. In this approach, meromorphic modular forms are sections over $\mathcal{M}^o(\mathrm{N})$, holomorphic modular forms are sections (of a canonically defined extension of $\underline{\omega}$) over $\mathcal{M}(\mathrm{N})$, and cusp forms of weight $k+2$ are sections of $\underline{\omega}^{\otimes k}\otimes \Omega^1$ over $\mathcal{M}(\mathrm{N})$, where Ω^1 denotes the sheaf of differentials on $\mathcal{M}(\mathrm{N})$.

I.2.2 Basic Properties

The basic properties of the modules of p-adic modular forms with growth conditions are explored in [Ka73]. We summarize in this section those which will be useful to us later.

First, we note that when the p is nilpotent in the ring B, one can determine completely the module of (meromorphic) p-adic modular forms from the classical spaces. We have:

Proposition I.2.3 *When p is nilpotent in B and N is prime to p, there is a canonical isomorphism*

$$\mathsf{F}(B,k,\mathrm{N};r) \cong \left(\bigoplus_{j=0}^{\infty} F(B,k+j(p-1),\mathrm{N})\right) / (\mathrm{E}_{p-1}-r)\ , \tag{I.1}$$

where $F(B,k,\mathrm{N})$ denotes the space of classical meromorphic modular forms of weight k and level N over B.

Proof: We give an idea of the proof for the case when $\mathrm{N} \geq 3$, in which all the functors in question are representable; the cases $\mathrm{N}=1,2$ follows by the usual methods.

The main point is that the scheme $\mathcal{M}^o(\mathrm{N})$ is *affine*, and that the functor "sections of $\underline{\omega}^{1-p}$ over $\mathcal{M}^o(\mathrm{N})$" is represented by the relatively affine (and hence affine) scheme

$$\underline{Spec}_{\mathcal{M}^o}(Symm(\underline{\omega}^{p-1}));$$

adding the condition that the section satisfy $Y \cdot \mathrm{E}_{p-1} = r$ gives the scheme

$$\underline{Spec}_{\mathcal{M}^o}(Symm(\underline{\omega}^{p-1})/(\mathrm{E}_{p-1}-r)),$$

and the Leray spectral sequence does the rest, since we are looking for global sections of certain sheaves. (This shows also that the formal scheme representing the functor "elliptic curves over p-adic rings with level structure $\imath$ and r-structure Y" is an affine formal scheme.) For details, see [Ka73, Prop. 2.3.1]. □

The analogous result for holomorphic forms is not true unless r is a p-adic unit, in which case we might as well assume $r=1$. Then we have:

Proposition I.2.4 *When p is nilpotent in B and N is prime to p, there is a canonical isomorphism*

$$\mathsf{M}(B,k,\mathrm{N};1) \cong \left(\bigoplus_{j=0}^{\infty} M(B,k+j(p-1),\mathrm{N})\right) / (\mathrm{E}_{p-1}-1)\ , \tag{I.2}$$

where $M(B,k,\mathrm{N})$ denotes the space of classical holomorphic modular forms of weight k and level N over B.

Proof: This is proved in passing (for $\mathrm{N} \geq 3$, when things are representable) in [Ka75a]. The point here is that the scheme obtained by deleting the supersingular points in $\mathcal{M}(\mathrm{N})$ is *affine* (because $\underline{\omega}$ has positive degree, and hence is ample), and the proof of the preceding proposition still works. When r is not a unit, the scheme obtained is the covering of $\mathcal{M}(\mathrm{N})$ given by

$$\underline{Spec}_{\mathcal{M}}(Symm(\underline{\omega}^{p-1})/(\mathrm{E}_{p-1} - r)),$$

which is *not* affine. (This will also follow from the results relating this space to the space of generalized p-adic modular functions which we will consider later.) □

Remark: It is worth noting that the above theorem makes sense even for the case $k < 0$, where we simply take the spaces of classical forms of negative weight to be 0. This is also the case for the following results.

When p is not nilpotent in B, one has a properly p-adic situation. In this case, one can give a description of $\mathsf{M}(B, k, \mathrm{N}; r)$ both as an inverse limit of classical objects and in terms of a "basis". First we have:

Proposition I.2.5 *Let $\mathrm{N} \geq 3$, and suppose that $k \neq 1$ or that $k = 1$ and $\mathrm{N} \leq 11$. Let B be any p-adically complete ring, and suppose that $r \in B$ is not a zero-divisor in B. Then we have an isomorphism*

$$\mathsf{M}(B, k, \mathrm{N}; r) \cong \varprojlim_{n} \mathrm{H}^0(\mathcal{M}(N), \bigoplus_{j=0}^{\infty} \underline{\omega}^{k+j(p-1)}) \otimes_{\mathbf{Z}_p} (\mathrm{B}/p^n B)/(\mathrm{E}_{p-1} - r) ,$$

where $\mathcal{M}(\mathrm{N}) = \overline{\mathcal{M}_1}(\mathrm{N})$ denotes the compactified moduli scheme over $\mathbf{Z}_p$ for elliptic curves with a $\Gamma_1(\mathrm{N})^{arith}$-structure and $\underline{\omega}$ denotes the invertible sheaf on $\mathcal{M}(\mathrm{N})$ obtained by pulling back the sheaf of differentials on the universal elliptic curve via the zero-section.

Proof: Once again, this is a matter of looking at the scheme classifying "test objects with growth condition r" and then using the Leray-Serre spectral sequence; see [Ka73, Thm. 2.5.1]. The restriction on the level when $k = 1$ is due to the fact that one does not have a base-change theorem for modular forms of weight 1 and level $\mathrm{N} \geq 12$ (see the discussion in [Ka73]). □

We now give a more interesting description, due to Katz, of the spaces $\mathsf{M}(B, k, \mathrm{N}; r)$; essentially, we show that one may choose a "Banach basis" for $\mathsf{M}(B, k, \mathrm{N}; r)$ in terms of classical modular forms. We first note that the map of spaces of classical modular forms

$$M(\mathbf{Z}_p, k + j(p-1), \mathrm{N}) \xrightarrow{\mathrm{E}_{p-1}} M(\mathbf{Z}_p, k + (j+1)(p-1), \mathrm{N})$$

given by multiplication by E_{p-1} admits a (non-canonical) section (this is [Ka73, Lemma 2.6.1], which again is a cohomological calculation). Choosing such sections once and for all for each $j \geq 0$ (a non-canonical procedure which should be thought of as analogous to a choice of basis) we get

$$A(\mathbf{Z}_p, k, j, \mathrm{N}) \subset M(\mathbf{Z}_p, k + j(p-1), \mathrm{N})$$

such that

$$M(\mathbf{Z}_p, k + j(p-1), \mathrm{N}) \cong \mathrm{E}_{p-1} \cdot M(\mathbf{Z}_p, k + (j-1)(p-1), \mathrm{N}) \oplus A(\mathbf{Z}_p, k, j, \mathrm{N}).$$

We also set

$$A(\mathbf{Z}_p, k, 0, \mathrm{N}) = M(\mathbf{Z}_p, k, \mathrm{N})$$

(which is just 0 when k is negative) and

$$A(B, k, j, \mathrm{N}) = A(\mathbf{Z}_p, k, j, \mathrm{N}) \otimes_{\mathbf{Z}_p} B.$$

Thus we have isomorphisms

$$\begin{array}{ccc} \displaystyle\bigoplus_{a=0}^{j} A(B, k, a, \mathrm{N}) & \xrightarrow{\ \sim\ } & M(B, k + j(p-1), \mathrm{N}) \\ \sum b_a & \longmapsto & \sum \mathrm{E}_{p-1}^{j-a} b_a. \end{array} \tag{I.3}$$

Now let $A^{rigid}(B, k, \mathrm{N})$ denote the B-module of all sums

$$\sum_{a=0}^{\infty} b_a, \qquad b_a \in A(B, k, a, \mathrm{N})$$

such that $b_a \to 0$ in the obvious sense, i.e., b_a becomes more and more divisible by p in $M(B, k + a(p-1), \mathrm{N})$ as $a \to \infty$. (Notice that $A^{rigid}(B, k, \mathrm{N})$ does not depend on r, as the notation suggests.) It is clear that $A^{rigid}(B, k, \mathrm{N})$ is naturally a p-adically complete B-module. If B is a p-adically complete discrete valuation ring with fraction field K, taking $A^{rigid}(B, k, \mathrm{N})$ as the "unit ball" defines a p-adic norm on $A^{rigid}(B, k, \mathrm{N}) \otimes K$ which makes this a p-adic Banach space over K.

The spaces of p-adic modular forms then turn out to be all isomorphic (but with different isomorphisms) to $A^{rigid}(B, k, \mathrm{N})$, via an "expansion in terms of the chosen basis".

Proposition I.2.6 *Suppose that either $k \neq 1$ or $\mathrm{N} \leq 11$. Then the inclusion, via (I.3), of $A^{rigid}(B, k, \mathrm{N})$ in the p-adic completion of*

$$\mathrm{H}^0(\mathcal{M}(\mathrm{N}), \bigoplus_{j \geq 0} \underline{\omega}^{k + j(p-1)})$$

induces, for any r, an isomorphism

$$\begin{array}{ccc} A^{rigid}(B, k, \mathrm{N}) & \xrightarrow{\ \sim\ } & \mathsf{M}(B, k, \mathrm{N}; r) \\ \sum b_a & \longmapsto & \displaystyle\sum \frac{r^a b_a}{\mathrm{E}_{p-1}^a}, \end{array} \tag{I.4}$$

where $\sum r^a b_a / \mathrm{E}_{p-1}^a$ *is the p-adic modular form with growth condition r defined by*

$$\left(\sum_{a\geq 0} \frac{r^a b^a}{\mathrm{E}_{p-1}^a}\right)(\mathrm{E}/A, \omega, \imath, Y) = \sum_{a\geq 0} b_a(\mathrm{E}/A, \omega, \imath)\cdot Y^a.$$

In particular, $\mathsf{M}(B, k, \mathrm{N}; r)$ *is a p-adically complete B-module.*

Remark: In other words, since we have chosen Y to satisfy

$$Y \cdot \mathrm{E}_{p-1}(\mathrm{E}, \omega) = r,$$

it makes sense to evaluate "$r^a \mathrm{E}_{p-1}^{-a}$" on curves with r-structures, and, since our ring is chosen to be p-adically complete, it makes sense to look at convergent sums of such modular forms. Since modular forms may be multiplied by E_{p-1}, one must use the splitting referred to above to ensure uniqueness of the expansion. The point of the theorem is that all p-adic modular forms with growth condition r are obtained in this way, so that this gives a "Banach basis" (in quotes, because the "basis coefficients" will be modular forms rather than numbers) for our space.

Proof: see [Ka73, Section 2.6]. □

The description we have just obtained has many useful corollaries, especially concerning the relation between the spaces of overconvergent forms when one varies the growth condition r.

Corollary I.2.7 *Let* $r_2 = r\cdot r_1$ *in B. Under the hypotheses above, the canonical mapping*

$$\mathsf{M}(B, k, \mathrm{N}; r_2) \longrightarrow \mathsf{M}(B, k, \mathrm{N}; r_1)$$

defined by transposition from the map of functors

$$(\mathrm{E}/A, \omega, \imath, Y) \longrightarrow (\mathrm{E}/A, \omega, \imath, r\cdot Y)$$

is injective, and is given in terms of the "basis" by

$$\begin{array}{ccc} A^{rigid}(B, k, \mathrm{N}) & \longrightarrow & A^{rigid}(B, k, \mathrm{N}) \\ \sum b_a & \longmapsto & \sum r^a b_a. \end{array} \tag{I.5}$$

This corollary allows us to identify the space of "overconvergent" p-adic modular forms $\mathsf{M}(B, k, \mathrm{N}; r)$ (where r is not a unit in B) as a subspace of the space $\mathsf{M}(B, k, \mathrm{N}; 1)$ of ("non-overconvergent", because defined only outside the supersingular disks) p-adic modular forms. The description we get amounts to saying that the overconvergent forms are those whose Laurent expansions around the omitted supersingular disks converge especially rapidly (the "b_a" should tend to zero "better than linearly"). To be precise:

Corollary I.2.8 *Under the previous hypotheses, and assuming r is not a unit in B, let $f \in \mathsf{M}(B,k,\mathrm{N};1)$, and let*

$$f = \sum_{a \geq 0} \frac{b_a}{\mathrm{E}_{p-1}^a}$$

be its expansion. Then, for any $m \geq 0$, $p^m f$ is in the image of $\mathsf{M}(B,k,\mathrm{N};r)$ if and only if r^a divides $p^m b_a$ in $M(B,k+a(p-1),\mathrm{N})$, for every $a \geq 0$, and $r^{-a}b_a \to 0$ in the same sense as above. If B is a discrete valuation ring, this is equivalent to $\mathrm{ord}_p(b_a) \geq a \cdot \mathrm{ord}_p(r) - m$ and $\mathrm{ord}_p(r^{-a}b_a) \to \infty$, where we normalize ord *by* $\mathrm{ord}_p(p) = 1$. *In particular, if K is the fraction field of B, then $f \in \mathsf{M}(B,k,\mathrm{N};r) \otimes K$ if and only if $r^{-a}b_a \to 0$ as $a \to \infty$.*

Proof: immediate from the previous corollary. □

Note in particular that

$$p^m f \in \mathsf{M}(B,k,\mathrm{N};r) \qquad \text{and} \qquad f \in \mathsf{M}(B,k,\mathrm{N};1)$$

does not imply $f \in \mathsf{M}(B,k,\mathrm{N};r)$, so that the natural p-adic topologies on the spaces $\mathsf{M}(B,k,\mathrm{N};r)\otimes K$ and $\mathsf{M}(B,k,\mathrm{N};1)\otimes K$ are distinct. (In this statement, we are of course assuming B is a domain and denoting its fraction field by K.) In fact,

$$(\mathsf{M}(B,k,\mathrm{N};r) \otimes K) \cap \mathsf{M}(B,k,\mathrm{N};1)$$

is *dense* in $\mathsf{M}(B,k,\mathrm{N};1)$, because it contains all *finite* sums of the form

$$\sum_{\text{finite}} b_a \mathrm{E}_{p-1}^{-a},$$

while, as the next result shows, $\mathsf{M}(B,k,\mathrm{N};r)$ is actually a "very small" subspace of $\mathsf{M}(B,k,\mathrm{N};1)$.

We give $\mathsf{M}(B,k,\mathrm{N};r) \otimes K$ the p-adic topology determined by making the B-submodule $\mathsf{M}(B,k,\mathrm{N};r)$ the closed unit disk; this is, so to speak, its "natural" p-adic topology: it is the p-adic topology which makes the map induced by Proposition I.2.6 a linear homeomorphism of topological B-modules. With this topology, $\mathsf{M}(B,k,\mathrm{N};r)\otimes K$ becomes a p-adic Banach space, and we have the following very significant result:

Corollary I.2.9 *Under the hypotheses above, assume that B is a discrete valuation ring, that B/pB is finite, and that $r_2 = r \cdot r_1$ in B, where r is not a unit in B. Let K be the field of fractions of B. Then the canonical map $\mathsf{M}(B,k,\mathrm{N};r_2)\otimes K \longrightarrow \mathsf{M}(B,k,\mathrm{N};r_1)\otimes K$ obtained as in the previous corollary is a completely continuous homomorphism of p-adic Banach spaces.*

Proof: One needs only check that the image of $\mathsf{M}(B,k,\mathrm{N};r_2)$ (the unit ball in $\mathsf{M}(B,k,\mathrm{N},r_2) \otimes K$) is relatively compact. Since it is contained in $\mathsf{M}(B,k,\mathrm{N};r_1)$, it is bounded, hence one need only check that its reduction modulo a power of p is a finite set, which is clear from the expression of the inclusion in terms of the "basis". □

The next important property of our construction has to do with comparing the p-adic properties of modular forms and of their q-expansions. As we remarked above, we can only expect these properties to correspond if we exclude the supersingular curves; in our setup, that amounts to taking $r = 1$. It turns out that this is also sufficient:

Proposition I.2.10 *With hypotheses as in Proposition I.2.5, let $x \in B$ be any element which divides some power of p. For any p-adic modular form $f \in \mathsf{M}(B,k,\mathrm{N};1)$, the following are equivalent:*

i. $f \in x \cdot \mathsf{M}(B,k,\mathrm{N};1)$,

ii. the q-expansion of f lies in $x \cdot B[[q]]$.

Proof: see [Ka73, Prop. 2.7]. □

Thus, the p-adic norm on $\mathsf{M}(B,k,\mathrm{N};1)$ is induced by the p-adic norm on $B[[q]]$. (Note that this is definitely not the case if r is not a unit in B; in that case, the topology induced by the q-expansion map is *weaker* than the "natural" topology. $\mathsf{M}(B,k,\mathrm{N};r)\otimes K$ is not complete with the q-expansion topology, since its image is dense in $\mathsf{M}(B,k,\mathrm{N};1)\otimes K$.) This shows that the spaces with growth condition 1 are the ones to consider in order to obtain information about congruences between q-expansions. For example, the following is an immediate consequence of Proposition I.2.10 and Corollary I.2.8:

Corollary I.2.11 *Let $f \in \mathsf{M}(B,k,\mathrm{N};r)$; then there exists a classical modular form $b_0 \in M(B,k,\mathrm{N})$ such that $f(q) \equiv b_0(q) \pmod p$.*

The point here is that b_0 has the same weight and level as f; on the other hand, it is perfectly possible that $b_0(q) \equiv 0 \pmod p$, in which case the result seems less interesting.

Since Serre's definition of p-adic modular forms is formulated in terms of limits (in the p-adic topology of $B[[q]]$) of q-expansions, one would guess, after Proposition I.2.10, that Serre's space will be related to our space $\mathsf{M}(B,k,\mathrm{N};1)$. This is in fact the case. The first step is to get another description of the space of p-adic modular forms with growth condition $r = 1$:

Proposition I.2.12 *Under the preceding hypotheses, given a power series $f(q) \in B[[q]]$, the following are equivalent:*

i. $f(q)$ is the q-expansion of a p-adic modular form $f \in \mathsf{M}(B,k,\mathrm{N};1)$,

ii. for every $n \geq 1$, there exists $m \geq 1$ such that $m \equiv 0 \pmod{p^{n-1}}$ and a classical modular form $g_n \in M(B,\mathrm{N},k+m(p-1))$ whose q-expansion is congruent to $f(q)$ modulo p^n in $B[[q]]$.

Proof: The main difficulty is to show that the reduction modulo p^n of the q-expansion of a p-adic modular form is *classical* (i.e., is the reduction of the q-expansion of a classical modular form); see [Ka73, Prop. 2.7.2]. □

Thus, the space of p-adic modular forms (of integral weight) defined by Serre in terms of limits of (q-expansions of) classical modular forms (see [Se73]) coincides with $\mathsf{M}(B, k, \mathrm{N}; 1)$. Of course, Serre also considers more general weights $\chi \in \mathrm{Hom}_{conts}(\mathbf{Z}_p^\times, \mathbf{Z}_p^\times)$. We will consider these later, since they appear naturally in the theory of generalized p-adic modular functions.

There are clearly relations between the various spaces of modular forms with growth condition when one varies the weight. The simplest of these comes from the obvious remark that E_{p-1} is "invertible", i.e., that $\mathrm{E}_{p-1}^{-1} \in \mathsf{M}(\mathbf{Z}_p, 1-p, 1; 1)$. Hence, we have

Corollary I.2.13 *Multiplication by* E_{p-1} *gives an isomorphism*

$$\mathsf{M}(B, k, \mathrm{N}; 1) \xrightarrow{\mathrm{E}_{p-1}} \mathsf{M}(B, k+p-1, \mathrm{N}; 1).$$

The corresponding map on the "bases" is given by

$$\begin{array}{ccc} A^{rigid}(B, k, \mathrm{N}) & \longrightarrow & A^{rigid}(B, k+p-1, \mathrm{N}) \\ (b_0, b_1, b_2, \ldots) & \longmapsto & (\mathrm{E}_{p-1}b_0 + b_1, b_2, \ldots). \end{array}$$

As to the image of the subspace of overconvergent forms, we have

$$\begin{array}{rcl} \mathrm{E}_{p-1}\mathsf{M}(B, k, \mathrm{N}; r) & \subset & \mathrm{E}_{p-1}M(B, k, \mathrm{N}) + r\mathsf{M}(B, k+p-1, \mathrm{N}; r) \\ & \subset & \mathsf{M}(B, k+p-1, \mathrm{N}; r). \end{array}$$

In particular, if $M(B, k, \mathrm{N}) = 0$ *(for example, if* $k < 0$*), we get an isomorphism*

$$\mathsf{M}(B, k, \mathrm{N}; r) \xrightarrow{\frac{1}{r}\mathrm{E}_{p-1}} \mathsf{M}(B, k+p-1, \mathrm{N}; r).$$

Proof: This is all immediate by considering the expansion in (I.4) and using Corollary I.2.8. □

To some extent, this result shows that the theory of modular forms of negative integral weight is determined by the theory for positive integral weight. However, the isomorphism is not equivariant for the action of the Hecke operators, which makes it less interesting from the point of view of the following chapters, where we will mostly be considering eigenforms under the Hecke operators.

I.3 Generalized p-adic Modular Functions

Generalized p-adic modular functions were first introduced by Katz in his papers [Ka75b], [Ka77] [Ka76], and [Ka75a]. They represent a generalization of what was done before, and, as we shall see, they contain all the spaces we have discussed so far. The ring of generalized p-adic modular functions is the ideal context for studying congruences between modular forms of different weights, and also for considering universal problems, as we shall do later with respect to Galois representations.

We begin by giving the definition and the basic properties of p-adic modular functions, and define the diamond operators which act on them. Then, using the action of the diamond operators, we define the weight and nebentypus of a p-adic modular functions and relate the definition to the classical one. In an appendix, we explain how Serre's "p-adic modular forms of weight χ" fit into the picture.

There are slight variations in approach among the several papers of Katz quoted above; our approach is closest to that in [Ka76], and we usually direct the reader there for further details, especially of proofs, most of which we only sketch. For an overview, the reader might also check the relevant sections of [Ka75b].

I.3.1 Definition

We will define generalized p-adic modular functions as functions on trivialized elliptic curves (see Section I.1). Recall, first, that we have defined a p-adic ring to be a ring B that is complete and separated in the p-adic topology, so that we have

$$B = \varprojlim_n B/p^n B.$$

We will define a p-adic modular function as something which takes values on trivialized elliptic curves defined over such rings. More precisely, the functor from the category of p-adic rings (with homomorphisms that are continuous in the p-adic topology) to the category of sets given by

$$\{p\text{-adic rings A}\} \longrightarrow \left\{\begin{array}{l}\text{isomorphism classes of triples} \\ (\mathrm{E}/A, \varphi, \imath) \text{ where E is an elliptic curve} \\ \text{over A, } \varphi \text{ is a trivialization, and } \imath \\ \text{is a compatible arithmetic level } \mathrm{N}p^\nu\text{-} \\ \text{structure}\end{array}\right\}$$

is representable by a p-adic ring $\mathbf{W}(\mathbf{Z}_p, \mathrm{N}p^\nu)$. For any p-adic ring B, the same functor restricted to B-algebras A is represented by $\mathbf{W}(B, \mathrm{N}p^\nu) = \mathbf{W}\hat{\otimes}B$.

To construct the ring $\mathbf{W}$, we first note that, since it is a p-adic ring, we must have

$$\mathbf{W} = \varprojlim_n \mathbf{W}/p^n\mathbf{W} = \varprojlim_n \mathbf{W}(\mathbf{Z}/p^n\mathbf{Z}, \mathrm{N}p^\nu),$$

so that we need only specify the rings $\mathbf{W}_n = \mathbf{W}(\mathbf{Z}/p^n\mathbf{Z}, \mathrm{N}p^\nu)$. For this, we recall that a trivialization may be thought of as a compatible family of $\Gamma_1(p^\mu)$-structures. Recall that $\mathcal{M}(\mathrm{N}p^\nu)$ denotes the moduli space of elliptic curves over $\mathbf{Z}_p$ with a $\Gamma_1(\mathrm{N}p^\nu)^{arith}$-structure (with the cusps added, but of course still affine if $\nu \geq 1$), and that $\mathcal{M}^o(\mathrm{N}p^\nu)$ is the subscheme obtained by deleting the cusps. For every $m \geq \nu$ let $\mathbf{W}_{n,m}$ denote the coordinate ring of the affine scheme $\mathcal{M}^o(\mathrm{N}p^m) \otimes \mathbf{Z}/p^n\mathbf{Z}$ (when $\nu = 0$ and $m = 0$, one must take the coordinate ring of the affine scheme obtained by deleting the supersingular points from $\mathcal{M}^o(\mathrm{N}) \otimes \mathbf{Z}/p^n\mathbf{Z}$). Then we set

$$\mathbf{W}_n = \varinjlim_m \mathbf{W}_{n,m}, \tag{I.6}$$

so that

$$\mathbf{W} = \varprojlim_n \varinjlim_m \mathbf{W}_{n,m}. \tag{I.7}$$

Note that this definition is independent of the exponent ν, so that we have

$$\mathbf{W}(\mathbf{Z}_p, \mathrm{N}p^\nu) = \mathbf{W}(\mathbf{Z}_p, \mathrm{N}).$$

This is also clear from the modular description of $\mathbf{W}$, since the trivialization φ and the requirement that the level structure be compatible with it determine $\Gamma_1(p^\nu)^{arith}$-structures for all $\nu > 0$. (For more details of the construction of $\mathbf{W}$, see [Ka75b], [Ka76], and [Ka77].) An element $f \in \mathbf{W}$ is called a *generalized p-adic modular function.*

Given a trivialized elliptic curve $(\mathrm{E}/_A, \varphi, \imath)$ over a p-adic ring A, and a generalized p-adic modular function $f \in \mathbf{W}$, we get a *value*

$$f(\mathrm{E}/_A, \varphi, \imath) \in A,$$

which depends only on the isomorphism class of the trivialized curve; this process commutes with base change of p-adically complete $\mathbf{Z}_p$-algebras. The modular function f is determined by all of its values (tautologically, since f *is* its value on the universal trivialized elliptic curve over $\mathbf{W}$).

We want to define when a p-adic modular function is holomorphic; once again, we do this by considering the Tate curve. Thus, let Tate(q) be the Tate elliptic curve over $\widehat{\mathbf{Z}_p((q))}$; there are canonical maps

$$\begin{array}{llcl} \varphi_{can}: & \widehat{\mathrm{Tate}(q)} & \xrightarrow{\ \sim\ } & \hat{\mathbf{G}}_m \\ \imath_{can}: & \mu_{\mathrm{N}p^\nu} & \hookrightarrow & \mathrm{Tate}(q)[\mathrm{N}p^\nu]. \end{array}$$

Then we can evaluate any $f \in \mathbf{W}$ at $(\mathrm{Tate}(q), \varphi_{can}, \imath_{can})$ to get an element $f(q) \in \widehat{\mathbf{Z}_p((q))}$, which we call the q-expansion of f. Mapping f to $f(q)$ gives a homomorphism

$$\mathbf{W} \xrightarrow{f \to f(q)} \widehat{\mathbf{Z}_p((q))},$$

which we call the q-expansion map. We will say that $f \in \mathbf{W}$ is *holomorphic* if $f(q) \in \mathbf{Z}_p[[q]]$ (which in fact implies that we have $f(\mathrm{Tate}(q), \varphi, \imath) \in \mathbf{Z}_p[[q]]$ for any level structure $\imath$ and any trivialization φ), and we denote by $\mathbf{V} = \mathbf{V}(\mathbf{Z}_p, \mathrm{N})$ the subring of $\mathbf{W}$ consisting of the holomorphic generalized p-adic modular functions. Finally, we extend these definitions to $\mathbf{W}(B, \mathrm{N})$ and $\mathbf{V}(B, \mathrm{N})$ for any p-adically complete $\mathbf{Z}_p$-algebra B in the obvious way (i.e., by restricting the functor to be represented to schemes over B, or equivalently by restricting our test objects to elliptic curves over B-algebras.)

The ring $\mathbf{V}$ of holomorphic modular functions may also be constructed directly by noting that the schemes $\mathcal{M}(\mathrm{N}p^m)\otimes\mathbf{Z}/p^n\mathbf{Z}$ are themselves affine when $m>0$, and taking $\mathbf{V}_{n,m}$ to be their coordinate rings, and by taking $\mathbf{V}_{n,0}$ to be the coordinate ring of the affine scheme $\mathcal{M}^{ord}(\mathrm{N})\otimes\mathbf{Z}/p^n\mathbf{Z}$ obtained by deleting the supersingular points in the scheme $\mathcal{M}(\mathrm{N})\otimes\mathbf{Z}/p^n\mathbf{Z}$. Then we get

$$\mathbf{V}=\varprojlim_{n}\varinjlim_{m}\mathbf{V}_{n,m}, \tag{I.8}$$

as before. This is the approach followed in [Ka75a]. The $\mathbf{V}_{n,m}$ are étale over $\mathbf{V}_{n,0}$, and may also be described in terms of the characters of the étale fundamental group of $\mathcal{M}^{ord}(\mathrm{N})\otimes\mathbf{Z}/p^n\mathbf{Z}$ defined by the action on the étale quotient of the kernel of p^n on the universal elliptic curve. For more details, see Katz's treatment in [Ka73, Chapter 4].

There is nothing special about the "weight zero" aspect of this construction (i.e., the fact that it uses the coordinate rings, thus the "forms of weight zero" on the various incomplete modular curves $\mathcal{M}(\mathrm{N}p^m)\otimes\mathbf{Z}/p^n\mathbf{Z}$). In fact, as Katz observes in [Ka75a], if $m\geq 1$, one can choose a nonvanishing section of $\underline{\omega}$ over each of the affine curves $\mathcal{M}(\mathrm{N}p^m)\otimes\mathbf{Z}/p^n\mathbf{Z}$, so that we have $\mathrm{H}^0(\mathcal{M}(\mathrm{N}p^m)\otimes\mathbf{Z}/p^n\mathbf{Z},\mathcal{O})\cong\mathrm{H}^0(\mathcal{M}(\mathrm{N}p^m)\otimes\mathbf{Z}/p^n\mathbf{Z},\underline{\omega}^{\otimes k})$ for any k. This allows us, so to speak, to think of the construction as involving forms of any weight k. In fact, as we shall see, the spaces of p-adic modular forms of weight k are all contained in $\mathbf{V}$.

Finally, we need to define the ideal of parabolic modular functions. For this, we define $\mathbf{V}_{n,m}^{cusp}$ to be the ideal of $\mathbf{V}_{n,m}$ determined by the requirement that the q-expansions at all the cusps (i.e., at the Tate curve with all possible arithmetic level $\mathrm{N}p^m$-structures) be in $q\cdot(\mathbf{Z}/p^n\mathbf{Z})[[q]]$, then define

$$\mathbf{V}_{par}=\varprojlim_{n}\varinjlim_{m}\mathbf{V}_{n,m}^{cusp}.$$

Equivalently, we can define, for $m\geq 1$,

$$\mathbf{V}_{n,m}^{cusp}=\omega^{\otimes -2}\mathrm{H}^0(\mathcal{M}(\mathrm{N}p^m)\otimes\mathbf{Z}/p^n\mathbf{Z},\,\Omega^1),$$

where Ω^1 is the sheaf of differentials and ω is the canonical nonvanishing section of $\underline{\omega}$ mentioned above. Note, then, that the analogous construction for a p-adic ring B yields $\mathbf{V}_{par}(B,\mathrm{N})=\mathbf{V}_{par}\hat{\otimes}B$, and that, by construction, $\mathbf{V}_{par}(B,\mathrm{N}p^\nu)$ is independent of ν.

I.3.2 The q-expansion map

The most fundamental result about the relation between p-adic modular functions and their q-expansions is what is known as the q-expansion principle. It captures the fact that by excluding the supersingular curves (via the requirement of a trivialization) we have made the p-adic properties of the modular functions correspond well to the p-adic properties of their q-expansions. It is this result which gives the q-expansion map its fundamental role in the theory.

Theorem I.3.1 *Let B be a p-adic ring. The q-expansion map*

$$\mathbf{W}(B,\mathrm{N}) \longrightarrow B\widehat{((q))}$$

is injective, and the cokernel

$$B\widehat{((q))}\big/\mathbf{W}(B,\mathrm{N}p^{\nu})$$

is flat over B.

Proof: The main point is to show that the q-expansion map is injective irrespective of the ring B, and hence in particular for B/pB, whence the theorem. This property is related to the irreducibility of the moduli space of trivialized elliptic curves; a proof taking this approach, but using the language of algebraic stacks, can be found in [Ka75b]. (See also the proof in [Ka75a], which uses a different language.) □

Another useful result is

Proposition I.3.2 *Let $B \subset B'$ be p-adic rings. Then we have a natural inclusion $\mathbf{W}(B,\mathrm{N}p^{\nu}) \subset \mathbf{W}(B',\mathrm{N}p^{\nu})$, satisfying: for $f \in \mathbf{W}(B',\mathrm{N}p^{\nu})$ we have*

$$f \in \mathbf{W}(B,\mathrm{N}p^{\nu}) \iff f(q) \in B\widehat{((q))}.$$

Proof: see [Ka76, Chapter 5]. □

This means that the ring over which a modular function is defined is determined by its q-expansion coefficients, and justifies the classical approach of starting with complex modular forms and then requiring that the q-expansion coefficients belong to various subrings.

These two results are together known as "the q-expansion principle", and are fundamental in all that follows. In intuitive terms, they mean that the situation "near the cusps" determines what happens at all ordinary curves (since we have omitted the supersingular disks). One should remark that they clearly remain true if we substitute $\mathbf{W}$ by $\mathbf{V}$ everywhere.

I.3.3 Diamond operators

The diamond operators are defined by varying the level structure and the trivialization of the given elliptic curve by the action of the natural groups. In terms of the action of these operators, we then define the weight and the nebentypus of a generalized p-adic modular function (when they exist!). Let $\mathbf{V} = \mathbf{V}(\mathbf{Z}_p,\mathrm{N})$ be the ring of holomorphic p-adic modular functions, as above. Since we are mainly interested in $\mathbf{V}$ (rather than $\mathbf{W}$), we define the diamond operators only for $\mathbf{V}$; it is clear however, that the same definition works in general.

Let

$$G(\mathrm{N}) = \mathbf{Z}_p^{\times} \times \left(\mathbf{Z}/\mathrm{N}\mathbf{Z}\right)^{\times}.$$

We define an action of $(x,y) \in G(\mathrm{N})$ on $\mathbf{V}$ by

$$\langle x,y\rangle f(\mathrm{E},\varphi,\imath) = f(\mathrm{E},x^{-1}\varphi,y\imath),$$

where y acts on $\imath$ by the canonical action of $\mathbf{Z}/\mathrm{N}\mathbf{Z}$ on μ_{N} and x^{-1} acts on φ via the $\mathbf{Z}_p$-action on $\Gamma_1(p^\infty)^{arith}$-structures derived from the action of $\mathbf{Z}/p^n\mathbf{Z}$ on μ_{p^n}.

(The definition, of course, thinks of $\mathbf{V}$ as $\mathbf{V}(\mathbf{Z}_p,\mathrm{N})$; if we think of $\mathbf{V}$ as $\mathbf{V}(\mathbf{Z}_p,\mathrm{N}p^\nu)$, then we must let $\langle x,y\rangle$ act on a $\Gamma_1(\mathrm{N}p^\nu)^{arith}$-structure $\imath$ by acting by y on μ_{N} and by x on μ_{p^ν}, preserving the compatibility between level structure and trivialization.)

There is a close connection between the diamond operators and the spaces $\mathbf{V}_{n,m}$ which we used to construct $\mathbf{V}$. Let $\Gamma = 1+p\mathbf{Z}_p \subset \mathbf{Z}_p^\times$ denote the subgroup of one-units, and let $\Gamma_i \subset \Gamma$ denote its unique subgroup of index p^i, and let $\mathbf{V}_{n,\infty} = \varinjlim_m \mathbf{V}_{n,m} = \mathbf{V}\otimes\mathbf{Z}/p^n\mathbf{Z}$. Then we have:

Proposition I.3.3 *The subring $\mathbf{V}_{n,m} \subset \mathbf{V}_{n,\infty}$ consists precisely of those elements of $\mathbf{V}_{n,\infty}$ which are fixed under the action of Γ_m via the diamond operators. In particular,*

$$\mathbf{V}_{1,1} = (\mathbf{V}\otimes\mathbf{Z}/p\mathbf{Z})^\Gamma = \{f \in \mathbf{V}\otimes\mathbf{Z}/p\mathbf{Z} \mid \langle\gamma,1\rangle f = f,\ \forall\gamma\in\Gamma\}.$$

Proof: See [Ka75a]. This result is intuitively clear, since the action of Γ_m on the trivialization does not change the arithmetic p^m-structure it determines, and since $\mathbf{V}_{n,m}$ is precisely the part of $\mathbf{V}_{n,\infty}$ that depends only on a level p^m-structure. The proof simply makes this precise. □

The diamond operators are ring homomorphisms, i.e., we have

$$\langle x,y\rangle(fg) = (\langle x,y\rangle f)(\langle x,y\rangle g).$$

Thus, we may decompose $\mathbf{V}$ in terms of the characters of any finite subgroup (of order prime to p, if we wish to avoid denominators) of $G(\mathrm{N})$, and it will sometimes be convenient to do so.

Remark: It is *not* possible to decompose $\mathbf{V}$ in terms of the characters of all of $G(\mathrm{N})$ (as Katz points out in [Ka75a]); in fact, the sum of the isotropic subspaces corresponding to the various characters of $G(\mathrm{N})$ is a proper subring of $\mathbf{V}$. To see this, consider first the special case of a character $\chi : \mathbf{Z}_p^\times \longrightarrow \mathbf{Z}_p^\times$ given by $\chi(x) = x^k$, for some integer k, and suppose that $f \in \mathbf{V}$ satisfies $\langle x,1\rangle f = \chi(x)f = x^k f$. Then, modulo p, f is invariant under the subgroup of one-units $\Gamma = 1+p\mathbf{Z}_p \subset \mathbf{Z}_p^\times$. It follows from Proposition I.3.3 that the reduction of f mod p belongs to $\mathbf{V}_{1,1}$, which is only a small part of the reduction of $\mathbf{V}$. Since any continuous character $\mathbf{Z}_p^\times \longrightarrow \mathbf{Z}_p^\times$ must map Γ into Γ, this is in fact true for any f on which the diamond operators act through such a character. Thus, the reduction modulo p of the sum of all the isotropic subspaces corresponding to characters $\mathbf{Z}_p^\times \longrightarrow \mathbf{Z}_p^\times$ must be contained in the proper subring $\mathbf{V}_{1,1}$ of $\mathbf{V}\otimes\mathbf{F}_p$.

It is easy to see that the diamond operators permute the trivializations on $\mathrm{Tate}(q)$ transitively, and stabilize the subspace $\mathbf{V}_{par}(\mathbf{Z}_p,\mathrm{N})$ of parabolic modular functions.

I.3.4 Weight and nebentypus

The diamond operators allow us to give a natural definition of the weight and nebentypus of a generalized p-adic modular function (when they exist). Given a continuous character $\chi : \mathbf{Z}_p^\times \longrightarrow B^\times$, we say that a generalized p-adic modular function $f \in \mathbf{V}(B,\mathrm{N})$ has weight χ (as a modular function of level N) if

$$\langle x,1\rangle f = \chi(x)f,$$

for all $x \in \mathbf{Z}_p$. If $\chi(x) = x^k$ for some $k \in \mathbf{Z}$, we say f is of weight k. In addition, if $k \in \mathbf{Z}$ and ε is a character of $\mathbf{Z}/\mathrm{N}\mathbf{Z}$, we say that f is of weight k and nebentypus ε (as a modular function of level N) if

$$\langle x,y\rangle f = x^k\varepsilon(y)f.$$

Finally, whenever there is a continuous character $\chi : G(\mathrm{N}) \longrightarrow B^\times$ giving the action of the diamond operators on a modular function f, we will say that χ is the weight-and-nebentypus character of f (because it contains the weight and the nebentypus "mixed together"; note that when p divides the order $\phi(\mathrm{N})$ of $(\mathbf{Z}/\mathrm{N}\mathbf{Z})^\times$ this can be a difficulty). As we shall see, these definitions turn out to be consistent with the usual definitions of weight and nebentypus (for modular forms of level N).

An important special case is that of forms of weight (i,k). In [Se73], Serre considers continuous characters $\chi \in X = \mathrm{Hom}(\mathbf{Z}_p^\times, \mathbf{Z}_p^\times)$. Decomposing $\mathbf{Z}_p^\times$ as $\mathbf{Z}_p^\times = \mathbf{Z}/(p-1)\mathbf{Z}\times\Gamma$, he shows that $X \cong \mathbf{Z}/(p-1)\mathbf{Z} \times \mathbf{Z}_p$, where, given $x = (u,s) \in \mathbf{Z}_p^\times = \mathbf{Z}/(p-1)\mathbf{Z} \times \Gamma$, an element $(i,k) \in \mathbf{Z}/(p-1)\mathbf{Z} \times \mathbf{Z}_p$ corresponds to the character $(u,s) \mapsto \chi_{(i,k)}(x) = u^i s^k \in \mathbf{Z}_p^\times$. The element $(1,0)$ corresponds then to the Teichmüller character ω, and we can rewrite the character corresponding to (i,k) as

$$x \longmapsto \chi_{(i,k)}(x) = (\omega(x))^i(\frac{x}{\omega(x)})^k;$$

when k is an integer, one can write $\chi_{(i,k)}(x) = (\omega(x))^{i-k}x^k$ for any $x \in \mathbf{Z}_p^\times$, and we will sometimes prefer to think in this way. We will say that a p-adic modular function $f \in \mathbf{V}$ is a of p-adic weight (i,k) if we have $\langle x,1\rangle f = \chi_{(i,k)}(x)f$. As we will see ahead, these are precisely Serre's p-adic modular forms of weight (i,k).

As a matter of general policy, one should reserve the expression "modular *form*" for modular functions which have weights, in contrast to more general modular *functions* $f \in \mathbf{V}$. The point of the next two sections is to relate this convention to the spaces of modular forms which we already know, both classical and p-adic.

I.3.5 Modular forms and modular functions

Given a trivialization φ on E/A, we can pull back the canonical invariant differential on $\hat{\mathbf{G}}_m$ to obtain an invariant differential on $\hat{\mathrm{E}}$, which then extends to an invariant differential on E. (If A is flat over $\mathbf{Z}_p$, φ is uniquely determined by the differential thus obtained, and one can characterize which differentials correspond to trivializations — see [Ka76, Section 5.4] .) This allows us to define maps from spaces of modular forms to $\mathbf{V}$(B,N).

Let $dt/(1+t)$ denote the canonical invariant differential on $\hat{\mathbf{G}}_m$; the map

$$(\mathrm{E},\varphi,\imath) \longrightarrow (\mathrm{E},\, \varphi^*\left(\frac{dt}{1+t}\right),\imath)$$

defines, for each k, a homomorphism of B-modules

$$\begin{array}{ccc} M(B,k,\mathrm{N}) & \longrightarrow & \mathbf{V}(B,\mathrm{N}) \\ f & \longmapsto & \tilde{f}, \end{array} \tag{I.9}$$

and hence a homomorphism of rings

$$\bigoplus_{k=0}^{\infty} M(B,k,\mathrm{N}) \longrightarrow \mathbf{V}(B,\mathrm{N}), \tag{I.10}$$

given by

$$\tilde{f}(\mathrm{E},\varphi,\imath) = f(\mathrm{E},\, \varphi^*\left(\frac{dt}{1+t}\right),\imath).$$

This preserves q-expansions, so that the map (I.9) is injective (by the q-expansion principle for classical modular forms). The map (I.10) is injective if and only if B is flat over $\mathbf{Z}_p$ (see [Ka77, 1.1]). It is clear that $f \in M(B,\mathrm{N},k)$ implies that $\tilde{f}$ is of weight k in the sense described above, and similarly for f of weight k and nebentypus ε, so that the diamond operators map the image of $M(B,k,\mathrm{N})$ in $\mathbf{V}$ to itself. Similarly, the image of the subspace $S(B,k,\mathrm{N})$ of cusp forms is contained in $\mathbf{V}_{par}$ and is mapped to itself by all the diamond operators.

The situation for p-adic modular forms is analogous. Any trivialized elliptic curve is fiber-by-fiber ordinary, so that $\mathrm{E}_{p-1}(\mathrm{E}/A,\omega)$ is invertible in A, for any invariant differential ω. Hence, the map

$$(\mathrm{E},\varphi,\imath) \longrightarrow (\mathrm{E},\varphi^*\left(\frac{dt}{1+t}\right),\imath,(\mathrm{E}_{p-1}(\mathrm{E},\varphi^*\left(\frac{dt}{1+t}\right)))^{-1})$$

defines maps

$$\begin{array}{ccc} \mathsf{M}(B,k,\mathrm{N};1) & \longrightarrow & \mathbf{V}(B,\mathrm{N}) \\ f & \longmapsto & \tilde{f} \end{array}. \tag{I.11}$$

It is again clear that this map preserves q-expansions, that it maps the cusp forms into $\mathbf{V}_{par}$, and that, for $f \in \mathsf{M}(B,k,\mathrm{N};1)$, the image $\tilde{f}$ has weight k in the sense defined above. In fact, the p-adic modular forms of weight k and growth condition $r=1$ defined above are precisely the generalized p-adic modular functions of weight k. More generally, for any character

$$\chi \in \mathrm{Hom}_{conts}(\mathbf{Z}_p^\times,\mathbf{Z}_p^\times),$$

we have defined generalized p-adic modular functions of weight χ, and these coincide with Serre's p-adic modular forms of weight χ (which are defined as limits of q-expansions). The precise result is:

Proposition I.3.4 *Let B be a $\mathbf{Z}_p$-flat p-adically complete ring, and let $\chi : \mathbf{Z}_p^\times \longrightarrow \mathbf{Z}_p^\times$ be a continuous character. Then the set of $f \in \mathbf{V}(B, \mathrm{N})$ of weight χ coincides with the set of "p-adic modular forms of weight χ defined over B" in the sense of Serre, i.e., with the set of $f \in \mathbf{V}(B, \mathrm{N})$ for which there exists a sequence f_n of classical modular forms defined over B, of weight k_n and level N, such that*

- $f_n(q) \to f(q)$ *in the p-adic topology of $B[[q]]$,*
- $\chi(x) \equiv x^{k_n} \pmod{p^n}$, *for all $x \in \mathbf{Z}_p$.*

Proof: This is [Ka75a, Prop. A.1.6]. It is clear that the limit of such a sequence will necessarily by a generalized p-adic modular function of weight χ. For the converse, one needs to construct a sequence of classical forms. The difficulty is only in showing that each of the approximations constructed is indeed a classical modular form, that is, that it can be computed on a test object $(\mathrm{E}, \omega, \imath)$ (without needing to assume the curve is ordinary, and without having to give a trivialization). □

Taking the special case where χ is of the form $\chi(x) = x^k$ for some $k \in \mathbf{Z}$, we get:

Proposition I.3.5 *Let B be a p-adically complete ring, flat over $\mathbf{Z}_p$. Then the map I.11 is an inclusion, and its image is precisely the set of elements of $\mathbf{V}(B, \mathrm{N})$ which are of weight k as defined above.*

Proof: This follows from the previous proposition together with Proposition I.2.12. □

Remark: One can only expect to find an approximation theorem such as Proposition I.3.4 for forms of weight

$$\chi : \mathbf{Z}_p^\times \longrightarrow B^\times$$

in the case when the character χ can itself be approximated by characters of the form $\chi(x) = x^k$, with $k \in \mathbf{Z}$. Put in other terms, χ must belong to the closure in $\mathrm{Hom}_{conts}(\mathbf{Z}_p^\times, B^\times)$ of the image of $\mathbf{Z}$ under the map sending k to the character χ_k such that $\chi_k(x) = x^k$. Since we are assuming $p \geq 5$, this closure is just $\mathrm{Hom}_{conts}(\mathbf{Z}_p^\times, \mathbf{Z}_p^\times)$, so that the proposition above is the best possible. This is what makes Serre's p-adic modular forms of weight $\chi = \chi_{(i,k)}$ of special interest. On the other hand, if we allow the approximating classical forms to have level $\mathrm{N}p^\nu$ with $\nu \gg 0$, we can approximate *any* element in $\mathbf{V}$; this is a recent result of Hida which we discuss later (see section III.3).

One can also define inclusions of spaces of classical modular forms of level $\mathrm{N}p^\nu$,

$$M(B, k, \mathrm{N}p^\nu) \hookrightarrow \mathbf{V}(B, \mathrm{N}p^\nu) = \mathbf{V}(B, \mathrm{N})$$

and

$$S(B, k, \mathrm{N}p^\nu) \hookrightarrow \mathbf{V}_{par}(B, \mathrm{N}p^\nu) = \mathbf{V}_{par}(B, \mathrm{N}),$$

in a completely analogous manner, using the fact that the trivialization determines an arithmetic $\Gamma_1(p^\nu)$-structure

$$\varphi^{-1}|_{\mu_{p^\nu}} : \mu_{p^\nu} \hookrightarrow \mathrm{E}.$$

However, if $f \in M(B,k,Np^\nu)$, it will in general *not* be true that $\tilde{f}$ is of weight k in $\mathbf{V}$ in the sense defined above; rather, we will have $\langle x,1\rangle\tilde{f} = x^k f$ only for $x \in \mathbf{Z}_p$ such that $x \equiv 1 (\mathrm{mod}\ p^\nu)$. For forms with nebentypus, one gets that $\tilde{f}$ is an eigenform for the diamond operators, hence has a weight χ which will depend on the classical weight and on the p-part of the nebentypus character. Precisely, we will have

$$\langle x,1\rangle f = \varepsilon(x) x^k f,$$

where ε is the "p-part" of the nebentypus character, giving the action of $(\mathbf{Z}/p^\nu\mathbf{Z})^\times$. In particular, the image of a modular form of level Np which has a nebentypus (or even a p-nebentypus, i.e., on which $\mathbf{Z}/(p-1)\mathbf{Z} \in \mathbf{Z}_p^\times$ acts through a character) will be a p-adic modular function of some weight χ.

An important special case is that of modular forms on $\Gamma_1(\mathrm{N}) \cap \Gamma_0(p^\nu)$, the space of which we denote by $M(B,k,\Gamma_1(\mathrm{N}) \cap \Gamma_0(p^\nu)) \subset M(B,k,\mathrm{N}p^\nu)$. This is the subspace where the p-part of the nebentypus character is trivial (i.e., the nebentypus character is "tame"). One can check easily, then, that for any $f \in M(B,k,\Gamma_1(\mathrm{N}) \cap \Gamma_0(p^\nu))$, the image $\tilde{f}$ *is* of weight k in $\mathbf{V}$, i.e.,

$$\langle x,1\rangle f = x^k f,$$

for any $x \in \mathbf{Z}_p^\times$ (see the formula above!), so that we get an inclusion

$$M(B,k,\Gamma_1(\mathrm{N}) \cap \Gamma_0(p^\nu)) \hookrightarrow \mathsf{M}(B,k,\mathrm{N};1).$$

This can be described in modular terms: recall that modular forms on $\Gamma_1(\mathrm{N}) \cap \Gamma_0(p^\nu)$ can be viewed as functions of quadruples $(\mathrm{E}/A, \omega, \imath, \mathrm{H})$, where $(\mathrm{E}/A, \omega, \imath)$ is a test object of level N and H is a finite flat subscheme of rank p^ν of E; then the inclusion map is given by

$$\tilde{f}(\mathrm{E},\varphi,\imath) = f(\mathrm{E},\varphi^*(\frac{dt}{1+t}),\imath,\varphi^{-1}(\mu_{p^\nu})).$$

The theory of the fundamental subgroup will show that this is in fact independent of the trivialization, as it must be. (It will also follow that $\tilde{f}$ is in fact *overconvergent.*) One immediate consequence is:

Corollary I.3.6 *Let $f \in M(B,k,\Gamma_1(\mathrm{N}) \cap \Gamma_0(p^\nu))$. Then, for some $j \equiv k \pmod{p-1}$, there exists $g \in M(B,j,\mathrm{N})$ such that $f(q) \equiv g(q) \pmod{p}$.*

Proof: We have just shown that $\tilde{f} \in \mathsf{M}(B,k,\mathrm{N};1)$, and we know that $f(q) = \tilde{f}(q)$. By Proposition I.2.4, $\tilde{f}$ can be written, modulo p, as a "polynomial in E_{p-1}^{-1}"; multiplying by a high power of E_{p-1} then gives a classical modular form g with the desired property. □

Remark: For $f \in M(B,k,\mathrm{N}p^\nu) = M(B,k,\Gamma_1(\mathrm{N}p^\nu))$, there is no analogous result unless we require that f have a nebentypus. If f does have a nebentypus, so that $(\mathbf{Z}/p^\nu\mathbf{Z})^\times$ acts through a character $\varepsilon = \omega^i\psi$, where ω is the Teichmüller character and ψ is of p-power order, then one can find a classical modular forms as described in the corollary, except that j will depend not only on the weight k, but also on the power i of the Teichmüller character appearing in the nebentypus; specifically, we will have $j \equiv i + k \pmod{p-1}$. We leave further elaboration of the theory of "modular forms mod p" when the level is itself divisible by p to the reader.

I.3.6 Divided congruences

The last property of the ring $\mathbf{V}$ which we wish to emphasize is that it contains a dense subring which can be described in terms of congruences of classical modular forms. This fact is used in [Ka75a] to determine such congruences.

Let B be a p-adically complete discrete valuation ring, and let K be its field of fractions. We define the module of divided congruences of weight less than or equal to k as

$$\mathsf{D}_k(B,\mathrm{N}p^\nu) = \mathsf{D}_k = \{f \in \bigoplus_{j=0}^{k} M(K,j,\mathrm{N}p^\nu) \mid f(q) \in B[[q]]\},$$

and then define the *ring of divided congruences* by

$$\mathsf{D}(B,\mathrm{N}p^\nu) = \mathsf{D} = \varinjlim_k \mathsf{D}_k.$$

Note that D is much larger than the direct sum of the classical spaces $M(B,k,\mathrm{N}p^\nu)$ of modular forms defined over B. In fact, whenever we have a congruence of q-expansions

$$\sum f_i(q) \equiv 0 \pmod{p^m},$$

we have that

$$\frac{1}{p^m}\sum f_i \in \mathsf{D}.$$

For example,

$$\frac{\mathrm{E}_{p-1}-1}{p} \in \mathsf{D}.$$

We claim there is an injection $\mathsf{D} \hookrightarrow \mathbf{V}(B,\mathrm{N})$. To see this, let $\pi \in B$ be a uniformizer, and let $f = \sum f_i \in \mathsf{D}$, where $f_i \in M(K,\mathrm{N}p^\nu,i)$. Then we have $f(q) \in B[[q]]$, and, for some n, $\pi^n f \in \sum M(B,i,\mathrm{N}p^\nu)$, hence $\widetilde{\pi^n f} \in \mathbf{V}$. Then, since $(\pi^n f)(q) = \pi^n f(q)$, $f(q)$ is a π-torsion element in the quotient $\widehat{B((q))}/\mathbf{V}$. By the flatness statement in the q-expansion principle (see Theorem I.3.1 above), it follows that there exists $\tilde{f} \in \mathbf{V}$ such that $\tilde{f}(q) = f(q)$. Hence we may define

$$\begin{array}{ccc} \mathsf{D} & \overset{\alpha}{\hookrightarrow} & \mathbf{V} \\ f & \longmapsto & \tilde{f} \end{array}. \tag{I.12}$$

Note that the injectivity follows at once from the equality of the q-expansions, since B is flat over $\mathbf{Z}_p$. Then we have:

Proposition I.3.7 *For any $\nu \geq 0$, the image of* $\mathsf{D}(B, \mathrm{N}p^\nu)$ *under the map α is dense in* $\mathbf{V}(B, \mathrm{N})$.

Proof: This result is the first step in Katz's determination of the higher congruences between modular forms in [Ka75a]. It is clearly enough to consider the case when $\nu = 0$ (since increasing ν only enlarges the subspace in question), and then to show that, after reducing modulo p, the resulting map $\alpha_1 = \alpha \bmod p$ is onto. One shows first the following important fact:

Lemma I.3.8 *The map α_1 sends* $\sum M(B, i, \mathrm{N})$ onto $\mathbf{V}_{1,1}$, *with kernel equal to the ideal generated by* $\mathrm{E}_{p-1} - 1$, *and hence gives an isomorphism*

$$\sum M(B, i, \mathrm{N}) / (\mathrm{E}_{p-1} - 1) \xrightarrow{\sim} \mathbf{V}_{1,1}$$

Thus, $\mathbf{V}_{1,1}$ is the same as the space of "modular forms mod p" considered by Serre and Swinnerton-Dyer. The proof of the lemma (which is found in [Ka75a]) involves reinterpreting $\mathbf{V}_{1,1}$ slightly and then using the basic theory of the Hasse invariant.

To complete the proof of the proposition, Katz then constructs a sequence of generators for the Artin-Schreier extensions $\mathbf{V}_{1,m} \longrightarrow \mathbf{V}_{1,m+1}$, all of which are "explicitly" given divided congruences of classical forms, proving what we want. □

In what follows we will identify D with its image in $\mathbf{V}$. The fact that $\mathbf{V}$ possesses a dense subspace which is a direct limit of B-modules of finite rank will be crucial in what follows. It is clear that the diamond operators on $\mathbf{V}$ preserve the ring D of divided congruences: if $f \in \mathsf{D} = \mathsf{D}(B, \mathrm{N})$ and we write $f = \sum f_j$ with $f_j \in M(K, \mathrm{N}, j)$, we have $\langle x, 1 \rangle f = \sum x^j f_j \in \mathsf{D}$. It is hard to see how one could prove directly that this action preserves congruences of q-expansions.

Remark: An important variation in the above should be noted. It is sometimes important to exclude the constants, and define

$$\mathsf{D}'_k(B, \mathrm{N}p^\nu) = \{ f \in \bigoplus_{j=1}^{k} M(K, j, \mathrm{N}p^\nu) \mid f(q) \in B[[q]] \}$$

and

$$\mathsf{D}'(B, \mathrm{N}p^\nu) = \varinjlim_{k} \mathsf{D}'_k(B, \mathrm{N}p^\nu).$$

The ideal of D thus obtained is still dense in $\mathbf{V}$, because 1 can be approximated by suitably chosen Eisenstein series; for example, we have

$$\lim_{n \to \infty} \mathrm{E}_{p-1}^{p^n} = 1.$$

We will make some use of this different approach when dealing with Hecke operators and duality theorems[1].

[1]By analogy, one might consider the spaces

$$\mathsf{D}_{(i)} = \{ f \in \bigoplus_{j \geq i} M(K, j, p^\nu) \mid f(q) \in B[[q]] \}.$$

We would also like to obtain similar results for the space of parabolic modular forms. For this, we let

$$\mathsf{S}^k(B,\mathrm{N}p^\nu) = \{f \in \bigoplus_{j=0}^{k} S(K,j,\mathrm{N}p^\nu) \mid f(q) \in B[[q]]\}$$

and

$$\mathsf{S}(B,\mathrm{N}p^\nu) = \varinjlim_k \mathsf{S}^k(B,\mathrm{N}p^\nu).$$

Then, in the same way as before, we get an inclusion

$$\mathsf{S}(B,\mathrm{N}p^\nu) \hookrightarrow \mathbf{V}_{par}(B,\mathrm{N}p^\nu) = \mathbf{V}_{par}(B,\mathrm{N}).$$

The previous results then suggest that the image of the inclusion must be dense in $\mathbf{V}_{par}(B,\mathrm{N})$. This is indeed the case.

Proposition I.3.9 *For any $\nu \geq 0$, the image of* $\mathsf{S}(B,\mathrm{N}p^\nu)$ *is dense in* $\mathbf{V}_{par}(B,\mathrm{N})$.

Proof: Let us write $\mathsf{S} = \mathsf{S}(B,\mathrm{N}p^\nu)$ and $\mathbf{V}_{par} = \mathbf{V}_{par}(B,\mathrm{N})$, leaving B and the level understood. Since increasing the level only makes the space S larger, we may (and will) assume that the level is N, so that $\mathsf{S} = \mathsf{S}(B,\mathrm{N})$. We have a commutative diagram,

$$\begin{array}{ccc} \mathsf{S} & \hookrightarrow & \mathbf{V}_{par} \\ \downarrow & & \downarrow \\ \mathsf{D} & \hookrightarrow & \mathbf{V} \end{array}$$

in which all the arrows are inclusions and the image of the second horizontal arrow is dense.

Note that $\mathsf{S} = \mathsf{D} \cap \mathbf{V}_{par}$; in fact, let $f \in \mathsf{D}$, and write $f = \sum f_i$, with f_i of weight i. Then we have, for any $x \in \mathbf{Z}_p^\times$, $\langle x,1\rangle f = \sum x^i f_i$. If we write the q-expansion of f_i as

$$f_i(q) = a_0(i) + \dots,$$

then $f \in \mathbf{V}_{par}$ implies

$$\sum x^i a_0(i) = 0,$$

for all $x \in \mathbf{Z}_p^\times$. It then follows at once that $a_0(i) = 0$ for all i. Similarly, for $\langle x,y\rangle f$, consider $\langle x_1 x,y\rangle f$ with $x_1 \in \mathbf{Z}_p^\times$, and it then follows that all the f_i are cusp forms, as desired.

To show that S is dense in $\mathbf{V}_{par}$, it is sufficient to check that the map $\mathsf{S} \longrightarrow \mathbf{V}_{par}$ is surjective modulo p. That is, we want to show that, given a parabolic modular function $f \in \mathbf{V}_{par}$, we may find a divided congruence *of cusp forms* $g \in \mathsf{S}$ such that $f \equiv g \pmod p$. For this, we use the explicit construction of $\mathbf{V}_{par} \otimes B/pB$ as a direct limit of submodules $\mathbf{V}_{n,m}^{cusp} \subset \mathbf{V}_{n,m}$, and follow point by point Katz's proof of Proposition I.3.7.

These are clearly contained in $\mathbf{V}$, and are they dense in $\mathbf{V}$ for every i, for similar reasons.

Consider first the graded ideal of the graded ring $\bigoplus M(B,k,\mathrm{N})$ of classical modular forms of level N consisting of the cusp forms,

$$\bigoplus_{k=0}^{\infty} S(B,k,\mathrm{N}) \subset \mathsf{S}.$$

We know, by Lemma I.3.8, that the image of the ring $\bigoplus M(B,k,\mathrm{N})$ in $\mathbf{V}\otimes B/pB$ is precisely the subring $\mathbf{V}_{1,1}$. It is then trivial to see that the image of $\bigoplus_{k=0}^{\infty} S(B,k,\mathrm{N})$ is precisely $\mathbf{V}_{1,1}^{cusp}$. (Only surjectivity is a problem. For that, decompose $f \in \mathbf{V}_{1,1}$ with respect to the action of $\mu_{p-1} \subset \mathbf{Z}_p^\times$, say $f = \sum f_i$. The argument above shows that each f_i is again parabolic; multiplying enough times by E_{p-1}, we get the weight in the range for which the relevant base change theorem applies, and the f_i may then be lifted to classical cusp forms of weight $k \equiv i \pmod{p-1}$, and the result follows.)

To conclude, recall that $\mathbf{V}_{1,n}$ is étale over $\mathbf{V}_{1,1}$, and that we have explicit generators (see [Ka75a]), so that we may write

$$\mathbf{V}_{1,n} = \mathbf{V}_{1,1}[\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_{n-1}],$$

with $d_j \in \mathsf{D}$, where tilde denotes the image in $\mathbf{V}_{1,\infty} = \mathbf{V}\otimes B/pB$. Identifying $\mathbf{V}_{1,n}^{cusp}$ with the module of Kahler differentials $\Omega_{\mathbf{V}_{1,n}}$ of $\mathbf{V}_{1,n}$ over B/pB (via multiplication by $\omega^{\otimes 2}$, where ω is the canonical section of $\underline{\omega}$, and noting that all the curves involved are affine), we have an exact sequence

$$\mathbf{V}_{1,1}^{cusp}\otimes \mathbf{V}_{1,n} \longrightarrow \mathbf{V}_{1,n}^{cusp} \longrightarrow \Omega_{\mathbf{V}_{1,n}/\mathbf{V}_{1,1}} \longrightarrow 0.$$

Since $\mathbf{V}_{1,n}$ is étale over $\mathbf{V}_{1,1}$, the third term is zero, and hence we have a surjection

$$\left(\bigoplus S(B,k,\mathrm{N})\right)[d_1, d_2, \ldots, d_{n-1}] \twoheadrightarrow \mathbf{V}_{1,n}^{cusp}.$$

Passing to the limit, and since S is an ideal of D, we get a surjection

$$\mathsf{S} \twoheadrightarrow \mathbf{V}_{1,\infty} = \mathbf{V}\otimes B/pB,$$

which proves the proposition. □

In what follows, we will work mostly in the ideal $\mathbf{V}_{par}$ of parabolic p-adic modular functions, since it has better duality properties with respect to the Hecke operators, which is the theme of the next chapter.

I.3.7 Appendix: p-adic modular forms of weight χ

We are now able to tie together all the aspects of the theory to give a coherent account of the theory of p-adic modular forms of weight χ. These were first defined by Serre in [Se73] as limits of q-expansions. We begin by recalling the definition.

Definition I.3.10 *Let $\chi \in \mathrm{Hom}_{cont}(\mathbf{Z}_p^\times, \mathbf{Z}_p^\times)$ be a continuous character. We say $f(q) \in B[[q]]$ is a Serre p-adic modular form of weight χ and level* N *defined over B if there exists a sequence of classical modular forms f_n of weight k_n, level* N, *and defined over B such that:*

i. $f_n(q) \to f(q)$ *in the* p*-adic topology of* $B[[q]]$,

ii. $\chi(x) \equiv x^{k_n} \pmod{p^n}$ *for all* $x \in \mathbf{Z}_p^\times$.

We have seen above that the continuous characters $\chi \in \mathrm{Hom}_{conts}(\mathbf{Z}_p^\times, \mathbf{Z}_p^\times)$ can be indexed by pairs $(i,k) \in (\mathbf{Z}/(p-1)\mathbf{Z}) \times \mathbf{Z}_p$, via the decomposition $\mathbf{Z}_p^\times = (\mathbf{Z}/(p-1)\mathbf{Z}) \times \Gamma$; recall that the correspondence is given by the formula

$$\chi_{(i,k)}(x) = \omega^i(x) \left(\frac{x}{\omega(x)}\right)^k,$$

where the second factor makes sense for any $k \in \mathbf{Z}_p$ because $x/\omega(x) \in \Gamma$ is a one-unit. Thus, it is clear that, for any $\chi = \chi_{(i,k)}$ as above, there exists a sequence k_n as in the definition above, so that

$$\chi_{(i,k)}(x) \equiv x^{k_n} \pmod{p^n}.$$

It is useful to note that this condition determines k_n modulo $p^{n-1}(p-1)$, and that we may chose the k_n to be increasing with n in the definition above (by multiplying the f_n by appropriate Eisenstein series).

Serre p-adic modular forms are, by definition, a kind of q-expansion; we have already obtained a modular interpretation, in Proposition I.3.4, which we repeat here:

Proposition I.3.11 *A series* $f(q) \in B[[q]]$ *is a Serre* p*-adic modular form of weight* $\chi = \chi_{(i,k)}$ *if and only if it is the* q*-expansion of a* p*-adic modular function* $f \in \mathbf{V}$ *which is of weight* χ*, that is, which satisfies the transformation law* $\langle x, 1\rangle f = \chi(x) f$*, for any* $x \in \mathbf{Z}_p^\times$.

We denote the space of p-adic modular forms of weight $\chi = \chi_{(i,k)}$ of level N defined over B by

$$\mathsf{M}(B, \chi, \mathrm{N}; 1) = \mathsf{M}(B, (i,k), \mathrm{N}; 1).$$

One sees immediately that, just as in the case of integral weight, we have

$$\mathsf{M}(B, \chi_{(i,k)}, \mathrm{N}; 1) = \varprojlim_n \mathsf{M}(B/p^nB, \chi_{(i,k)}, \mathrm{N}; 1) = \varprojlim_n \mathsf{M}(B/p^nB, k_n, \mathrm{N}; 1),$$

where the k_n form an approximating sequence to the character χ, as above. This shows that we have indeed obtained a modular interpretation, i.e., that p-adic modular forms of weight $\chi_{(i,k)}$ can be evaluated on elliptic curves with differential and level structure defined over a p-adic ring (by evaluating the reduction modulo p^n on the reduction of the given curve, and then taking the limit). The author does not know if this is true for p-adic modular functions which are of weight $\chi \in \mathrm{Hom}_{conts}(\mathbf{Z}_p^\times, B^\times)$, which cannot necessarily be approximated by classical modular forms as above.

Since we do know when a modular form of integral weight is overconvergent, the expression of $\mathsf{M}(B, (i,k), \mathrm{N}; 1)$ as an inverse limit allows us to define overconvergent p-adic modular forms of weight (i,k):

Definition I.3.12 *For any $r \in B$ and any character*

$$\chi_{(i,k)} \in \mathrm{Hom}_{conts}(\mathbf{Z}_p^\times, \mathbf{Z}_p^\times),$$

we define the space of p-adic modular forms of weight $\chi_{(i,k)}$ and growth condition r by:

$$\mathsf{M}(B, \chi_{(i,k)}, \mathrm{N}; r) = \varprojlim_n \mathsf{M}(B/p^n B, k_n, \mathrm{N}; r),$$

where k_n is a sequence of integers satisfying $\chi_{(i,k)}(x) \equiv x^{k_n} \pmod{p^n}$.

Since for every n there are maps

$$\mathsf{M}(B/p^n B, k_n, \mathrm{N}; r) \longrightarrow \mathsf{M}(B/p^n B, k_n, \mathrm{N}; 1),$$

taking the inverse limit gives a map

$$\mathsf{M}(B, \chi_{(i,k)}, \mathrm{N}; r) \longrightarrow \mathsf{M}(B, \chi_{(i,k)}, \mathrm{N}; 1).$$

It is *not* clear that this map is an inclusion, because the maps modulo p^n are not injective. (In the case of integral weight, we showed the injectivity as a consequence of the existence of the expansion (I.4) and of the description of the map in terms of that expansion.) This suggests the following two questions, which we have not been able to settle:

Question I.1 *Let the spaces $\mathsf{M}(B, \chi_{(i,k)}, \mathrm{N}; r)$ and the maps*

$$\mathsf{M}(B, \chi_{(i,k)}, \mathrm{N}; r) \xrightarrow{\alpha} \mathsf{M}(B, \chi_{(i,k)}, \mathrm{N}; 1)$$

be defined as above. Are the maps α inclusions? In other words, can we think of overconvergent forms of weight (i,k) (as defined above) as a certain kind of p-adic modular forms of weight (i,k)?

Question I.2 *Is there an analogue of the expansion (I.4) for forms of weight (i,k)? If so, does it provide a criterion for deciding if a given form is overconvergent?*

A negative answer to the first question would be very surprising; in fact, it would indicate that our definition is wrong.

A clue as to what is true is given by the fact that one may use Eisenstein series to relate different spaces of modular forms of weight $\chi_{(i,k)}$ and growth condition 1 to each other. In [Se73], Serre has constructed Eisenstein series $\mathrm{E}^*_{(0,j)}$, of p-adic weight $(0,j)$, satisfying $\mathrm{E}^*_{(0,j)} \equiv 1 \pmod{p}$. Then multiplication by $\mathrm{E}^*_{(0,k-i)}$ gives an isomorphism

$$\mathsf{M}(B, i, \mathrm{N}; 1) \xrightarrow{\sim} \mathsf{M}(B, \chi_{(i,k)}, \mathrm{N}; 1)$$

(which does not commute with the Hecke action to be defined in the next chapter). In the special case in which k is a positive integer, the Eisenstein series $\mathrm{E}^*_{(0,k-i)}$ is actually a *classical* modular form of weight $k - i$, level $\mathrm{N}p$, and nebentypus ω^{i-k}, so that the isomorphism maps the space of modular forms of weight i on $\Gamma_1(\mathrm{N}) \cap \Gamma_0(p)$ to the space of modular forms of weight k, level $\mathrm{N}p$, and nebentypus ω^{i-k} (which is precisely the space of modular forms of level $\mathrm{N}p$ which have p-adic weight (i, k)). A similar statement could be made for forms of level $\mathrm{N}p^{\nu}$ with the appropriate nebentypus characters, so that we may say that the isomorphism we have obtained preserves the classical subspaces (in the case when $k \in \mathbf{Z}$). It is not immediately clear what is the image of the space of overconvergent forms under this isomorphism, the difficulty being to decide whether the Eisenstein series in question is "overconvergent" i.e., whether it can be evaluated at a "not too supersingular" curve. This question, however, seems more accessible than the preceding ones.

The central role played by overconvergence in the spectral theory of the U operator (see the next chapter) makes it very interesting to obtain an analogue of the theory for all of the space $\mathbf{V}$, i.e., to define "overconvergent p-adic modular functions" as a certain subspace of $\mathbf{V}$ and extend the other aspects of the theory (especially the corollary measuring the size of the image of the overconvergent spaces in the full space). Answering the questions raised in this section would be a step in that direction.

Remark (for specialists): There is one subtle distinction between the theories of Serre and Katz that we have deliberately avoided above, having to do with the definition of a classical modular form over a finite field k. For Katz, such a thing is a function on elliptic curves over k (plus extra structure), or, equivalently, a section of an invertible sheaf defined over the modular curve corresponding to the situation (base-changed to k). Serre, on the other hand, *defines* a modular form over k to be the reduction of a classical modular form over a discrete valuation ring with residue field k. These definitions are known to be equivalent, *except* in the case when the weight is 1 and the level is greater than or equal to 12, in which case one simply does not know. This is the reason for restricting our results in the case of weight one. Note, however, that the difficulty disappears if we consider all weights together, since it is easy to see that any modular form of weight 1 over k will always have the same q-expansion as the reduction of some modular form of weight $p = 1 + (p - 1)$ (just multiply by E_{p-1} and note that the reduction map for modular forms of weight greater than one *is* onto).