Chapter II

The Hecke and U Operators

In this chapter we define p-adic versions of the classical Hecke operators. For the operators T_ℓ with $\ell \neq p$, this is quite easy, and may be done either by imitating the classical definition in terms of subgroups of order ℓ of elliptic curves or by simple extending the classical operators by continuity. The remaining case is more interesting. We define the Frobenius endomorphism, which corresponds to the "V_p" operator of Atkin and Lehner, and show that we may obtain a p-adic version of the U operator essentially as its trace. We then study the properties of these operators, especially with regard to their action on the spaces of modular forms with growth conditions, and study the spectral theory of the U operator acting on the spaces of p-adic modular forms with integral weight. This produces a number of interesting results, notably that U is a completely continuous operator on the spaces $\mathsf{M}(B, k, \mathrm{N}; r) \otimes K$ of overconvergent modular forms. As a consequence, once sees that, apart from the kernel of U, there are few overconvergent eigenforms for U. This allows us, for example to consider the characteristic power series of the U operator (which turns out not to depend on r). This connects nicely with the results obtained recently by Hida in the case when the eigenvalue is a unit (which implies overconvergence, as we will point out). By contrast, if we relax the requirement of overconvergence, we obtain a very large number of eigenforms for U, giving the theory a completely different aspect in that case.

II.1 Hecke Operators

The goal of this section is to define Hecke operators T_ℓ on the ring $\mathbf{V}$ of generalized p-adic modular functions. One can do this either by giving an intrinsic modular definition or by noticing that one can define Hecke operators on D from the classical Hecke operators by an inverse limit procedure. Since each of these methods gives important information about the Hecke operators, we will sketch both. In general, the inverse limit definition is more useful whenever we want to pass from classical results to results about p-adic modular functions, while the modular definition is better when we want to study questions of overconvergence.

II.1.1 Direct definition

Let E be an elliptic curve over a p-adic ring A, φ be a trivialization, and $\imath$ a $\Gamma_1(\mathrm{N}p^{\nu})^{arith}$-structure on E. Let ℓ be a rational prime, $\ell \neq p$, not dividing N. For any subgroup H of order ℓ in E, we can consider the quotient curve E/H; let π denote the canonical projection

$$\mathrm{E} \xrightarrow{\ \pi\ } \mathrm{E}/\mathrm{H},$$

and let $\check{\pi}$ denote the dual isogeny. Since ℓ does not divide $\mathrm{N}p^{\nu}$, both π and $\check{\pi}$ induce isomorphisms between the kernels of multiplication by $\mathrm{N}p^{\nu}$in E and in the quotient curve, so that we may define a level structure $\imath'$ on E/H by $\imath' = \check{\pi}^{-1} \circ \imath$.

$$\begin{tikzcd}
 & \mathrm{E} \\
\mu_{\mathrm{N}} \arrow[ur, "\imath"] \arrow[dr, "\imath'"'] & \\
 & \mathrm{E}/\mathrm{H} \arrow[uu, "\check{\pi}"']
\end{tikzcd}$$

In the same way, we can define a trivialization $\varphi' = \varphi \circ \pi^{-1}$ (this makes sense because π induces an isomorphism on the formal group over $\mathbf{Z}_p$, since $(\ell, p) = 1$).

$$\begin{tikzcd}
\hat{\mathrm{E}} \arrow[dd, "\pi"'] \arrow[dr, "\varphi"] & \\
 & \hat{\mathbf{G}}_m \\
\widehat{\mathrm{E}/\mathrm{H}} \arrow[ur, "\varphi'"'] &
\end{tikzcd}$$

Then we define, for $f \in \mathbf{V}$,

$$(\mathrm{T}_{\ell} f)(\mathrm{E}, \varphi, \imath) = \frac{1}{\ell} \sum_{\substack{\mathrm{H} \hookrightarrow \mathrm{E} \\ \#\mathrm{H} = \ell}} f(\mathrm{E}/\mathrm{H}, \varphi', \imath').$$

One can determine the effect of the T_{ℓ} on q-expansions by computing directly in terms of the Tate curve. If we take $f \in \mathbf{V}$, and assume that $f(q) = \sum a_n q^n$, one gets that

$$(\mathrm{T}_{\ell} f)(q) = \sum a_{n\ell} q^n + \frac{1}{\ell}((\langle \ell, \ell \rangle f)(q^{\ell}),$$

where $(\langle \ell, \ell \rangle f)(q^{\ell})$ denotes the image of the q-expansion of $\langle \ell, \ell \rangle f$ under the base change $q \mapsto q^{\ell}$, as in the classical case. If f has weight $k \in \mathbf{Z}$ and nebentypus ϵ, we have $\langle \ell, \ell \rangle f = \epsilon(\ell) \ell^k f$, so that we get

$$(\mathrm{T}_{\ell} f)(q) = \sum a_{n\ell} q^n + \epsilon(\ell) \ell^{k-1} f(q^{\ell}),$$

which is exactly the classical formula for the Hecke operators. Thus, we have indeed extended the classical Hecke operators to act on p-adic modular functions.

Since the Hecke and the diamond operators clearly commute, the Hecke operators preserve the space of forms of weight k, so that we get operators on $\mathsf{M}(B,k,\mathrm{N};1)$; to check that these preserve also the spaces of overconvergent forms, it is best to define them directly. So, for $f \in \mathsf{M}(B,k,\mathrm{N};r)$, we define $\mathrm{T}_\ell f$ by

$$(\mathrm{T}_\ell f)(E,\omega,\imath,Y) = \ell^{k-1} \sum_{\substack{\mathrm{H}\hookrightarrow\mathrm{E} \\ \#\mathrm{H}=\ell}} f(\mathrm{E}/\mathrm{H}, \check{\pi}^*\omega, \imath', \check{\pi}^*Y),$$

where $\pi : \mathrm{E} \longrightarrow \mathrm{E}/\mathrm{H}$ is the canonical projection, $\check{\pi}$ is the dual isogeny, and $\imath'$ is as above. One can then check immediately that this coincides with the operator induced by T_ℓ as defined above.

To define Hecke operators T_ℓ with ℓ dividing N (but different from p), we follow an analogous procedure, but sum only over those subgroups of order ℓ not contained in the image in E of the given level N structure. One checks immediately that the given level structure on E induces canonically a level structure the quotient by such subgroups. We thus obtain operators T_ℓ on $\mathbf{V}$ for $\ell|\mathrm{N}$, whose effect on q-expansions corresponds to that of the "U_ℓ" operators of Atkin-Lehner theory: if $f = \sum a_n q^n$, then

$$(\mathrm{T}_\ell f)(q) = \sum a_{\ell n} q^n.$$

Finally, we would like to define an operator corresponding to the prime p, which should extend the classical U operator (and not the classical T_p). Doing this in modular terms is slightly more subtle than the preceding; we shall do it later in this chapter, defining U in terms of the trace of an operator Frob which extends the classical V operator of Atkin-Lehner theory. This will allow us to see how U acts on the spaces of overconvergent forms. Simply to define the U operator on the full ring $\mathbf{V}$ is much simpler, and will be done in the next section by the inverse limit procedure explained below. On q-expansions, this acts as expected: if $f(q) = \sum a_n q^n$, then

$$(\mathrm{U}f)(q) = \sum a_{np} q^n.$$

We define the Hecke algebra $\mathbf{T}$ of $\mathbf{V}$ to be the completion of the commutative subalgebra of the space of $\mathbf{Z}_p$-linear endomorphisms of $\mathbf{V}$ generated by the T_ℓ (for all $\ell \neq p$), by the U operator, and by the diamond operators, where we give $\mathrm{End}_{\mathbf{Z}_p}(\mathbf{V})$ the compact-open topology. The action of the diamond operators then makes $\mathbf{T}$ an algebra over the profinite group ring $\mathbf{Z}_p[[G(\mathrm{N})]]$, and in particular over the algebras $\mathbf{\Lambda} = \mathbf{Z}_p[[\mathbf{Z}_p^\times]]$ and its subalgebra determined by the action of the one-units, the Iwasawa algebra $\Lambda = \mathbf{Z}_p[[\Gamma]]$. Since the Hecke operators defined above clearly preserve the ideal of parabolic forms, we may define an associated Hecke algebra as above; it is a quotient of $\mathbf{T}$ (via the restriction map), and we denote it by $\mathbf{T}_0$.

As in the classical case, the inclusion of the operators T_ℓ with ℓ dividing N and of U complicates the structure of the Hecke algebra (making it non-semisimple). Thus, as in the classical situation, we will sometimes wish to consider a restricted Hecke algebra $\mathbf{T}^\star$, which will be the completion of the algebra of endomorphisms of $\mathbf{V}$ generated by the diamond operators and the T_ℓ with ℓ not dividing $\mathrm{N}p$. It is thus a closed subalgebra of $\mathbf{T}$. The analogous restricted Hecke algebra for the space of parabolic p-adic modular functions will be denoted $\mathbf{T}_0^\star$. As we will see, it is essential to consider the full Hecke algebra to obtain duality theorems, but also essential to consider only the restricted Hecke algebra when studying the Galois representations attached to modular forms.

II.1.2 Hecke operators on divided congruences

In this section, we show that the action of the Hecke operators on $\mathbf{V}$ may also be obtained in terms of the dense submodule $\mathsf{D}' = \mathsf{D}'(B, \mathrm{N}p^\nu)$ of divided congruences of modular forms of level $\mathrm{N}p^\nu$ (where, for this construction, we must assume $\nu \geq 1$, in order to have an action of U). We first note that, the operators T_ℓ for $\ell \neq p$ and U act on classical modular forms of weight $j \geq 1$ and level $\mathrm{N}p^\nu$ and preserve congruences of modular forms over $\mathbf{Z}_p$, so that they act on the spaces D'_k. We define the Hecke algebra of D'_k, denoted by $\mathcal{H}'_k$, as the $\mathbf{Z}_p$-algebra of endomorphisms of D'_k generated by the endomorphisms induced by the T_ℓ, by U, and by the diamond operators. As before, we also define $\mathcal{H}'^\star_k$, by excluding the operator U and the T_ℓ with $\ell|\mathrm{N}$. Then it is clear that we have restriction maps

$$\mathcal{H}'_k \longrightarrow \mathcal{H}'_j$$

and

$$\mathcal{H}'^\star_k \longrightarrow \mathcal{H}'^\star_j$$

whenever $j < k$, and the inverse limits

$$\mathbf{T} = \varprojlim_k \mathcal{H}'_k$$

and

$$\mathbf{T}^\star = \varprojlim_k \mathcal{H}'^\star_k$$

are p-adically complete algebras of continuous endomorphisms of D', which are uniformly continuous in the q-expansion topology (i.e., the topology induced from $\mathbf{V}$). Since D' is dense in $\mathbf{V}$, the actions of $\mathbf{T}$ and $\mathbf{T}^\star$ extend to $\mathbf{V}$, and the action thus obtained coincides with the one we defined before (because it does so on the dense subspace D'—check on q-expansions). Thus, we can obtain the Hecke operators on $\mathbf{V}$ directly from the classical definition.

Remark: There is no special reason for using the space D' rather than D, other than the fact that later, when we consider the duality between spaces of modular functions and their corresponding Hecke algebras, we will need to exclude the constants. The construction here, of course, gives the same result whether we use D or D'.

Obtaining the Frob operator on $\mathbf{V}$ (which corresponds to the classical V operator) is a little more difficult because it does not preserve the level when acting on the classical spaces. Still, one need only consider the operator

$$\mathrm{Frob} : \mathrm{D}'(B, \mathrm{N}p^\nu) \longrightarrow \mathrm{D}'(B, \mathrm{N}p^{\nu+1}) \hookrightarrow \mathbf{V},$$

take the limit, and extend by continuity to an endomorphism of $\mathbf{V}$. In the following sections, we show how to interpret both this endomorphism and the U operator in modular terms, by giving an intrinsic definition, which avoids the description of $\mathbf{V}$ as the p-adic completion of an direct limit of classical spaces.

In the same way, by taking S instead of D$'$, we may define the parabolic Hecke algebra $\mathbf{T}_0$ and its restricted version $\mathbf{T}_0^\star$ as the inverse limit of the classical Hecke algebras on the spaces of divided congruences of cusp forms. In what follows, we will mainly be working with these algebras rather than with $\mathbf{T}$.

Since $\mathbf{T}_0(B, \mathrm{N}p^\nu)$ is defined to be the closure (in the compact-open topology) of the algebra of endomorphisms of $\mathbf{V}_{par}(B, \mathrm{N}p^\nu)$ generated by the Hecke and diamond operators, and since $\mathbf{V}_{par}(B, \mathrm{N}p^\nu) = \mathbf{V}_{par}(B, \mathrm{N})$, we have $\mathbf{T}_0(B, \mathrm{N}p^\nu) = \mathbf{T}_0(B, \mathrm{N})$, and of course similarly for $\mathbf{T}$ (if one thinks of the Hecke algebras as inverse limits, this should be taken as the *definition* of $\mathbf{T}_0(B, \mathrm{N})$, because we want an action of U and not of T_p). Note that the definition of $\mathbf{T}$, $\mathbf{T}^\star$, $\mathbf{T}_0$, and $\mathbf{T}_0^\star$ as inverse limits of compact $\mathbf{Z}_p$-algebras implies that they are *compact* topological $\mathbf{Z}_p$-algebras when given the inverse limit topology. We will show in Chapter III that this topology can be defined intrinsically (rather than in terms of the representation as an inverse limit); in fact, it turns out to be precisely the compact-open topology we considered in the preceding section. The reason this is the correct topology to consider (rather than, say the p-adic topology) is that it is the one for which the classical duality between modular forms and Hecke operators can be extended to the p-adic situation, as we will see in the next chapter.

II.2 The Frobenius Operator

In the theory of classical modular forms, one considers an operator, usually denoted "V", whose effect on q-expansions is

$$\sum a_n q^n \mapsto \sum a_n q^{np}.$$

This maps classical modular forms of level N to modular forms of level Np. Since modular forms of level Np^ν are p-adically of level N, it is reasonable to attempt to define a p-adic version of this operator, which should then map the ring $\mathbf{V}$ to itself. Following Katz, we call this the Frobenius operator, and denote it by Frob. It is an endomorphism of $\mathbf{V}$, and its existence is in some sense characteristic of the p-adic theory, in the sense that only p-adically does it preserve the level. As we shall see, it is quite interesting to analyze its action on the spaces of modular forms with growth condition.

We define the Frobenius operator on the ring $\mathbf{V}$ by using the fact that the trivialization determines a canonical subgroup of order p in E.

Definition II.2.1 *Let* $(E/_A, \varphi, \imath)$ *be a trivialized elliptic curve with an arithmetic level* N *structure. The fundamental subgroup of* E *is the A-sub-group scheme* $H \subset E$ *which extends the subgroup* $\varphi^{-1}(\mu_p)$ *of the formal completion of* E.

Thus, given a trivialized curve $(E, \varphi, \imath)$, we can consider the quotient E/H of E by its fundamental subgroup. Let π denote the canonical projection. Since p does not divide N, π and the dual isogeny $\check{\pi}$ induce isomorphisms between the kernel of multiplication by N in E and in E/H, so that we may define a level N structure on E/H by $\imath' = \check{\pi}^{-1} \circ \imath$. The picture is:

$$\begin{tikzcd} & E \\ \mu_N \arrow[ur, "\imath"] \arrow[dr, "\imath'"'] & \\ & E/H \arrow[uu, "\check{\pi}"'] \end{tikzcd}$$

Furthermore, the dual isogeny $\check{\pi}$ is *étale*, and hence induces an isomorphism on the formal completions, so that we may define a trivialization φ' of E/H by $\varphi' = \varphi \circ \check{\pi}$. Thus:

$$\begin{tikzcd} \hat{E} \arrow[dr, "\varphi"] & \\ & \hat{\mathbf{G}}_m \\ \widehat{E/H} \arrow[uu, "\check{\pi}"] \arrow[ur, "\varphi'"'] & \end{tikzcd}$$

We have then defined a map of functors

$$(E, \varphi, \imath) \longrightarrow (E/H, \varphi', \imath'),$$

which defines a map, the *Frobenius endomorphism* of $\mathbf{W}$,

$$\text{Frob} : \mathbf{W} \longrightarrow \mathbf{W},$$

by transposition: for $f \in \mathbf{W}$,

$$(\text{Frob}\, f)(E, \varphi, \imath) = f(E/H, \varphi', \imath').$$

An easy computation with the Tate curve shows that, on q-expansions, we have

$$(\text{Frob}\, f)(q) = f(q^p),$$

where "$f(q^p)$" denotes the image of $f(q) \in B\widehat{((q))}$ under the map $q \mapsto q^p$.

Some of the properties of Frob follow immediately from the definition and the effect on q-expansions. For example, the fact that Frob acts via $q \mapsto q^p$ on q-expansions implies that the Frobenius endomorphism Frob : $\mathbf{W} \longrightarrow \mathbf{W}$ generalizes the "V_p operator" of classical Atkin-Lehner theory; in particular, if $f \in M(B,k,\mathrm{N}) \subset \mathbf{V}$, we have $\mathrm{Frob}(f) \in M(B,k,\Gamma_1(\mathrm{N}) \cap \Gamma_0(p)) \subset M(B,k,\mathrm{N}p)$, and analogously for higher levels. It also follows at once that Frob preserves the subrings $\mathbf{V}$ and $\mathbf{V}_{par}$. Finally, note that when $B = \mathbf{F}_p$ we have $\mathrm{Frob}(f) = f^p$ (this also follows from the base change properties of modular functions), so that when $B = \mathbf{Z}_p$ we have a lifting of the p-power endomorphism of $\mathbf{V} \otimes \mathbf{F}_p$.

It is clear from the definition that the endomorphism Frob commutes with the diamond operators, and hence that it preserves weights. Therefore, Frob defines an endomorphism of $\mathsf{M}(B,k,\mathrm{N};1)$ for every integer k, and in fact of $\mathsf{M}(B,\chi,\mathrm{N};1)$ for any character $\chi : \mathbf{Z}_p^\times \longrightarrow \mathbf{Z}_p^\times$. It is not clear, however, what effect it has on overconvergent forms. Given the way we have defined it, this amounts to asking whether one can compute $\mathrm{Frob} f$ on a supersingular curve, and hence whether there is still a fundamental subgroup when the curve is "not too supersingular". It turns our that such a subgroup does exist for some supersingular curves. This is a result due to Lubin, which we now quote. To make the statements of the following theorems simpler, we here think of modular forms as sections of line bundles, so that Y should be thought of as a section of $\underline{\omega}^{\otimes(1-p)}$. Then we have:

Theorem II.2.2 *Let B be a complete discrete valuation ring with residue characteristic p and generic characteristic* 0, *and let $r \in B$ satisfy* $\mathrm{ord}(r) < p/(p+1)$, *where* ord *is normalized by* $\mathrm{ord}(p) = 1$. *To every test object* $(\mathrm{E}/A, \imath, Y)$ *of level* N *and growth condition r we may attach a finite flat subgroup scheme* $\mathrm{H} \subset \mathrm{E}$ *of rank p, called the fundamental subgroup of* E, *satisfying:*

- H *depends only on the isomorphism class of* $(\mathrm{E}/A, Y)$,
- *the formation of* H *commutes with arbitrary base change of p-adically complete B-algebras,*
- *if $p/r = 0$ in A,* H *is the kernel of the Frobenius map* $\mathrm{E} \longrightarrow \mathrm{E}^{(p)}$,
- *if* E/A *is the Tate curve* $\mathrm{Tate(q)}$ *over* $(A/p^nA)((q))$, *the fundamental subgroup* H *is the image of the canonical inclusion* $\mu_p \hookrightarrow \mathrm{Tate(q)}$.

Proof: The proof is a step-by-step construction of the fundamental subgroup, first as a formal subscheme of the formal group of E, which one then shows is a subgroup and extends to E. This requires a delicate analysis of the structure of the formal group of the curve. See [Ka73, Thm. 3.1], where both this and the following result are attributed to Lubin. □

Given that the fundamental subgroup exists, one must still check that it is possible to give the quotient curve an r-structure, i.e., one must check whether the quotient curve is more or less supersingular than the one we started with. It turns out that it is only possible to give the quotient curve an r^p-structure, so that the valuation of a lifting of the Hasse invariant is multiplied by p in the passage from the curve to its quotient by the fundamental subgroup.

Theorem II.2.3 *Under the previous hypotheses on B, suppose that* $\mathrm{ord}(r) < 1/(p+1)$. *Then there is one and only one way to attach to a test object* $(\mathrm{E}/A, \imath, Y)$ *of level* N *and growth condition r a test object* $(\mathrm{E}'/A, \imath', Y')$ *of level* N *and growth condition r^p, where*

$$\begin{cases} \mathrm{E}' = \mathrm{E}/\mathrm{H} \\ \imath' = \check{\pi}^{-1} \circ \imath \\ Y' \cdot \mathrm{E}_{p-1}(\mathrm{E}'/A, \imath') = r^p \end{cases}$$

such that

- *Y' depends only on the isomorphism class of* $(\mathrm{E}/A, Y)$,
- *the formation of Y' commutes with arbitrary base change of p-adically complete B-algebras,*
- *if $p/r = 0$ in A, then Y' is the inverse image $Y^{(p)}$ of Y on* $\mathrm{E}^{(p)} = \mathrm{E}'$.

Proof: See [Ka73, Thm. 3.2]. □

Since the quotient by fundamental subgroup is defined under the hypotheses of the theorems above, we can define the Frobenius homomorphism; however, because the quotient curve is more supersingular than the initial one, it does not follow that the homomorphism we obtain defines an endomorphism of the space of overconvergent forms. In fact, what we have is the following:

Theorem II.2.4 *Suppose* $\mathrm{N} \geq 3$, $p \nmid \mathrm{N}$, *and that either* $k \neq 1$ *or* $\mathrm{N} \leq 11$. *Let* $r \in A$ *satisfy* $\mathrm{ord}(r) < 1/(p+1)$. *For any* $f \in \mathsf{M}(B, k, \mathrm{N}; r^p)$, *the element* $\mathrm{Frob}(f) \in \mathsf{M}(B, k, \mathrm{N}; 1)$ *defined by*

$$\mathrm{Frob}(f)(\mathrm{E}/A, \omega, \imath, Y) = f(\mathrm{E}'/A, \check{\pi}^*(\omega), \imath', r^p \cdot Y'),$$

(with the notation of the previous theorem) satisfies

$$(\mathrm{Frob}\, f)(q) = f(q^p),$$

so that the map thus defined coincides with that induced by the Frobenius endomorphism of **W**. *Furthermore, we have*

$$\mathrm{Frob}(f) \cdot (\mathrm{E}_{p-1})^k \in \mathsf{M}(B, pk, \mathrm{N}; r),$$

or, equivalently,

$$r^k\, \mathrm{Frob}(f) \in \mathsf{M}(B, k, \mathrm{N}; r).$$

Proof: For all but the last statement, see [Ka73, Thm. 3.3]; the equivalence of the last two statements is clear by Corollary I.2.8. The decrease in overconvergence (from r^p to r) is, of course, due to the fact that the quotient curve is more supersingular; the non-integrality is due to the fact that the pull-back of a non-vanishing differential along the quotient map (or along the dual isogeny) is not non-vanishing (because both isogenies are of degree p). □

Thus, the Frobenius endomorphism defined above preserves the space of p-adic modular forms with growth condition 1, but, except in the case of weight zero, maps overconvergent forms to (less) overconvergent forms only up to multiplication by a power of r. In particular, if K denotes the field of fractions of B, Frob is a bounded linear homomorphism of p-adic Banach spaces from $\mathsf{M}(B,k,\mathrm{N};r^p)\otimes K$ to $\mathsf{M}(B,k,\mathrm{N};r)\otimes K$, but does *not* map $\mathsf{M}(B,k,\mathrm{N};r^p)$ to $\mathsf{M}(B,k,\mathrm{N};r)$, unless $k=0$.

The fact that the fundamental subgroup is defined allows to say a little more about the inclusion of classical forms into the space of p-adic modular forms discussed in Section I.3.5.

Corollary II.2.5 *Suppose* $\mathrm{N}\geq 3$, $p\nmid \mathrm{N}$. *Let* $f\in M(B,k,\Gamma_1(\mathrm{N})\cap\Gamma_0(p))$, *and let* $\tilde{f}$ *be its image in* $\mathsf{M}(B,k,\mathrm{N};1)$. *Suppose that* $r\in B$ *satisfies* $\mathrm{ord}(r)<p/(p+1)$. *Then we have*

$$\tilde{f}\in\mathsf{M}(B,k,\mathrm{N};r).$$

In other words, the inclusion

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

factors through the subspace $\mathsf{M}(B,k,\mathrm{N};r)$, *for any* $r\in B$ *of sufficiently small valuation.*

Proof: Simply define

$$\tilde{f}(\mathrm{E},\omega,\imath,Y)=f(\mathrm{E},\omega,\imath,\mathrm{H}),$$

where H is the fundamental subgroup. □

Remarks: *1)* This is false if $\mathrm{N}<3$ (put together Corollary I.2.11 with the classical q-expansion principle!).

2) In particular, it follows that if we have

$$f\in M(B,k,\mathrm{N})\subset\mathsf{M}(B,k,\mathrm{N};r)\qquad\text{and}\qquad\mathrm{ord}(r)<p/(p+1),$$

then $\mathrm{Frob}(f)\in\mathsf{M}(B,k,\mathrm{N};r)$, since classical forms of level N, which are clearly overconvergent, are mapped to classical forms on $\Gamma_1(\mathrm{N})\cap\Gamma_0(p)$, which are overconvergent to some degree (measured by the inequality on $\mathrm{ord}(r)$). As we have seen, this is *not* true for all overconvergent forms.

Of course, we also have inclusions

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

as described in Section I.3.5; given the preceding results, it is natural to ask whether the image is contained in some space of overconvergent forms. For this, all one needs is to define fundamental subgroups of order p^ν for every ν. For the case of trivialized curves, this is of course immediate:

Definition II.2.6 *Let $(\mathrm{E}/A,\varphi,\imath)$ be a trivialized elliptic curve with an arithmetic level* N *structure. The fundamental subgroup of order p^ν of* E *is the A-sub-group scheme $\mathrm{H}_\nu\subset\mathrm{E}$ which extends the subgroup $\varphi^{-1}(\mu_{p^\nu})$ of the formal completion of* E.

This is, in fact, exactly the subgroup we used to define the inclusion of $M(B,k,\Gamma_1(\mathrm{N})\cap\Gamma_0(p^\nu))$ in $\mathbf{V}$. For ordinary curves, there is also essentially no difficulty. Let F be the Frobenius map $\mathrm{E}\longrightarrow\mathrm{E}^{(p)}$ in characteristic p; the kernel of F^n has an étale dual, which can therefore be lifted uniquely, and we take the fundamental subgroup of order p^n to be the dual of this unique lifting. To obtain the result on overconvergence, we need fundamental subgroups for "not too supersingular" curves; for this, we simply iterate the construction of the fundamental subgroup of order p.

Theorem II.2.7 *Let B be a complete discrete valuation ring with residue characteristic p and generic characteristic* 0, *and let $r\in B$ have*

$$\mathrm{ord}(r)<\frac{1}{p^{\nu-2}(p+1)},$$

where ord *is normalized by* $\mathrm{ord}(p)=1$. *To every test object $(\mathrm{E}/A,\imath,Y)$ of level* N *and growth condition r we may attach a finite flat subgroup scheme $\mathrm{H}_\nu\subset\mathrm{E}$ of rank p^ν, called the fundamental subgroup of order p^ν of* E, *satisfying:*

- H_ν *depends only on the isomorphism class of* $(\mathrm{E}/A,Y)$,
- *the formation of H_ν commutes with arbitrary base change of p-adically complete B-algebras,*
- *if $p\cdot r^{-p^{n-1}}=0$ in A, H_ν is the kernel of the ν^{th} iterate of the Frobenius map $\mathrm{E}\xrightarrow{F^\nu}\mathrm{E}^{(p^\nu)}$,*
- *if E/A is the Tate curve* Tate(q) *over $(A/p^nA)((q))$, the fundamental subgroup H_ν is the image of the canonical inclusion $\mu_{p^\nu}\hookrightarrow$* Tate(q).

Proof: We use induction on ν. The case $\nu=1$ is precisely Theorem II.2.2. For $\nu\geq 2$, assume that we are given a test object $(\mathrm{E}/A,\omega,\imath,Y)$, with growth condition r such that

$$\mathrm{ord}(r)<\frac{1}{p^{\nu-2}(p+1)}\leq\frac{1}{p+1}.$$

By Theorem II.2.2, E has a fundamental subgroup H_1; we consider the quotient $E' = E/H_1$. By Theorem II.2.3, we get a test object $(E', \omega', \imath', Y')$, with growth condition r^p. Since

$$\mathrm{ord}(r^p) = p\ \mathrm{ord}(r) < \frac{1}{p^{\nu-3}(p+1)},$$

we get, by induction, a fundamental subgroup $H' \subset E'$ of order $\nu - 1$. Let $E'' = E'/H'$, and let f_ν be the composite map

$$f_\nu : E \longrightarrow E' \longrightarrow E''.$$

Then define $H_\nu = \ker(f_\nu)$, which clearly has the required properties (by induction from the case $\nu = 1$). One should note that, in the case $r = 1$, the fundamental subgroup is simply the dual of the unique lifting of the *étale* dual of the kernel of the iterated Frobenius map, or equivalently, the kernel of multiplication by p^ν in the formal group of E. □

We may read this theorem as saying that an *ordinary* curve over $\mathbf{Z}_p$ comes canonically equipped with a coherent sequence of subgroups of order p^ν (one might call this a $\Gamma_0(p^\infty)$-structure). For a (lifting of a) *supersingular* curve, a *part* of that sequence might still exist, depending in some sense on "how supersingular" the curve is, the measure being given by the valuation of $E_{p-1}(E, \omega)$ (which coincides with the valuation of any other lifting of the Hasse invariant in the range in question; it is interesting to note that the sequence of fundamental subgroups disappears completely exactly when $\mathrm{ord}(E_{p-1})$ is close to 1, which is also the point at which the p-adic valuation will begin to depend on the choice of the lifting). Dividing by the fundamental subgroup of order p makes the curve "more supersingular", i.e., increases $\mathrm{ord}(E_{p-1})$, and correspondingly shortens the sequence of fundamental subgroups.

Corollary II.2.8 *Let* $N \geq 3$, $p \nmid N$. *The canonical inclusion*

$$M(B, k, Np^\nu) \hookrightarrow \mathbf{V}$$

induces inclusions

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

for any r satisfying $\mathrm{ord}(r) < 1/p^{\nu-2}(p+1)$.

Proof: Set $\tilde{f}(E, \omega, \imath, Y) = f(E, \omega, \imath, H_\nu)$. □

In particular, we have shown that any classical modular form (i.e., any element of $M(B, k, Np^\nu)$ for some positive $k \in \mathbf{Z}$ and some ν) which is p-adically of *integral* weight (i.e., is in $\mathsf{M}(B, k, N; 1)$) is an *overconvergent* p-adic modular form, because having p-adic weight k implies that it belongs to

$$M(B, k, \Gamma_1(N) \cap \Gamma_0(p^\nu));$$

of course, the degree of overconvergence will decrease as ν increases. Note in particular that Frob is integral on $\mathsf{M}(B,k,\mathrm{N};1)$ and maps classical modular forms on $\Gamma_1(\mathrm{N})\cap\Gamma_0(p^\nu)$ to classical modular forms on $\Gamma_1(\mathrm{N})\cap\Gamma_0(p^{\nu+1})$, so that, if $f\in\mathsf{M}(B,k,\mathrm{N};r)$ is classical (hence, since the p-part of its nebentypus must be trivial, is a modular form on $\Gamma_1(\mathrm{N})\cap\Gamma_0(p^\nu)$ for some ν), Frob f will be overconvergent (*without* multiplying by a power of p, but for a different r):

$$\begin{array}{ccc} \mathsf{M}(B,k,\mathrm{N};r^p)\otimes K & \xrightarrow{\text{Frob}} & \mathsf{M}(B,k,\mathrm{N};r)\otimes K \\ \cup & & \cup \\ M(B,k,\Gamma_1(\mathrm{N})\cap\Gamma_0(p^\nu)) & \xrightarrow{\text{Frob}} & M(B,k,\Gamma_1(\mathrm{N})\cap\Gamma_0(p^{\nu+1})) \end{array}$$

where we must have $\mathrm{ord}(r)<1/p^{\nu-1}(p+1)$, as above, to make the inclusions true.

Our final result in this chapter is what will allow us later to give the modular definition of the U operator; it extends a result of Katz in [Ka73] from $\mathsf{M}(B,0,\mathrm{N};1)$ to $\mathbf{W}$.

Proposition II.2.9 *Let B be a p-adically complete discrete valuation ring, and let* $\mathbf{W}=\mathbf{W}(B,\mathrm{N})$. *Then the endomorphism* Frob : $\mathbf{W}\longrightarrow\mathbf{W}$ *is locally free of rank p.*

Proof: It is clearly enough to prove the theorem for $B=\mathbf{Z}_p$, since the general result then follows by base change. It is also clear, since $\mathbf{W}$ is p-adically complete, that the result for $\mathbf{W}=\mathbf{W}(\mathbf{Z}_p,\mathrm{N})$ follows from the analogous result for $\mathbf{W}/p\mathbf{W}=\mathbf{W}(\mathbf{F}_p,\mathrm{N})=\mathbf{W}_1$ (in the notation of Section I.3). Thus, we want to prove that Frob : $\mathbf{W}_1\longrightarrow\mathbf{W}_1$ is locally free of rank p. We may suppose $\mathrm{N}\geq 3$, since the cases of $\mathrm{N}=1,2$ will then follow by looking at fixed subrings under the action of the appropriate finite groups.

Recall that $\mathbf{W}_1=\varinjlim_m \mathbf{W}_{1,m}$, where $\mathbf{W}_{1,m}$ is the coordinate ring of the (affine) moduli scheme $\mathcal{M}^o(\mathrm{N}p^m)\otimes\mathbf{F}_p$ of elliptic curves over $\mathbf{F}_p$ with a $\Gamma_1(\mathrm{N}p^m)^{arith}$-structure, when $m>0$, and $\mathbf{W}_{1,0}$ is the scheme obtained by deleting the supersingular points from $\mathcal{M}^o(\mathrm{N})\otimes\mathbf{F}_p$.

As noted above, the Frobenius endomorphism induces $f\mapsto f^p$ on $\mathbf{W}_1$, and hence on each $\mathbf{W}_{1,m}$. Since the p-power morphism is locally free of rank p on the coordinate ring of any affine curve, Frob : $\mathbf{W}_{1,m}\longrightarrow\mathbf{W}_{1,m}$ is locally free of rank p. Hence we have commutative squares

$$\begin{array}{ccc} \mathbf{W}_{1,m+1} & \xrightarrow{\text{Frob}} & \mathbf{W}_{1,m+1} \\ \downarrow & & \downarrow \\ \mathbf{W}_{1,m} & \xrightarrow{\text{Frob}} & \mathbf{W}_{1,m} \end{array}$$

in which the horizontal arrows are locally free of rank p. The vertical arrows, on the other hand, are known to be étale of rank p when $m\geq 1$ (in fact, they are Artin-Schreier extensions—see [Ka73]). By counting ranks it follows that the squares are cartesian, and hence (see, for example, [EGA, IV.8.2]) that we can pass to the limit to get that Frob : $\mathbf{W}_1\longrightarrow\mathbf{W}_1$ is locally free of rank p, as desired. □

In the following section, we will use this result to define the U operator in terms of the trace of the Frobenius endomorphism.

II.3 The U Operator

In this section we define the U operator on generalized p-adic modular functions, show that it preserves the spaces of p-adic modular forms of weight k, and study its action on overconvergent forms. To begin with, we show that, after tensoring with the field of fractions K, the U operator in fact *improves* overconvergence. We then show that U induces a bounded linear endomorphism of the p-adic Banach space of modular forms with growth condition r, for any r such that $\operatorname{ord}(r) < p/(p+1)$. For r in the appropriate range, one even has that U is "almost integral", in the sense that its norm is equal to one in a topology equivalent to the standard p-adic topology.

We then go on to consider eigenforms for the U operator. The shape of the theory then depends in a fundamental way on whether we look at the full space of p-adic modular forms of weight k or at the spaces of overconvergent forms. On the full space, we show that one can produce an infinite number of eigenforms with eigenvalue λ for *every* element λ in the maximal ideal of B. By contrast, the fact that the U operator improves overconvergence implies that it is a completely continuous operator on the p-adic Banach space of overconvergent modular forms (in the sense of Serre; see [Se62]). One can then study its spectral theory. This allows us to define "slope α eigenspaces" for U which generalize (the integral weight case of) Hida's space of "ordinary p-adic modular forms". This will also show that there are few eigenforms for U outside its kernel, in the precise sense that if we fix the weight of f and the valuation of λ, one gets only a finite dimensional space of overconvergent forms of the given weight with eigenvalues of the given valuation. In contrast, it is clear that, even in the overconvergent case, $\ker(\mathrm{U})$ is quite large (in fact, infinite-dimensional), because of the Frobenius endomorphism: given *any* $f \in \mathrm{M}(B,k,\mathrm{N};r)$, we have $f - \mathrm{Frob}\,\mathrm{U}f \in \ker(\mathrm{U})$.

To get a complete generalization of Hida's theory, one would need to extend the spectral theory to the full space $\mathbf{V} \otimes K$; given our results about non-overconvergent eigenforms, such an extension is clearly impossible. (From this point of view, Hida's theory turns on the fact that ordinary eigenforms are necessarily overconvergent.) It might be possible, however, to extend his results to a dense subspace of $\mathbf{V} \otimes K$ consisting of "overconvergent" modular functions. The possibility of obtaining such a theory seems to be related to the question of how the theory for overconvergent p-adic modular forms of weight k varies with the weight. In the last part of this section, we obtain some preliminary results and make some conjectures as to what should be the case.

II.3.1 Definition

To define the U operator, we start with the Frobenius endomorphism $\mathrm{Frob} : \mathbf{V} \longrightarrow \mathbf{V}$, which, as was shown in the last section, is locally free of rank p. Therefore, there exists a trace homomorphism

$$Tr_{\mathrm{Frob}} : \mathbf{V} \longrightarrow \mathbf{V},$$

defined by

$$(Tr_{\mathrm{Frob}} f)(\mathrm{E}/A, \varphi, \imath) = \sum_{(\mathrm{E}_1, \varphi_1, \imath_1) \to (\mathrm{E}, \varphi, \imath)} f(\mathrm{E}_1, \varphi_1, \imath_1),$$

where the sum is taken over the triples $(\mathrm{E}_1, \varphi_1, \imath_1)$ which map (by quotient by the fundamental subgroup) to the given triple $(\mathrm{E}, \varphi, \imath)$. An easy calculation (essentially done in [Ka73, pp.22–23]) then shows that if $f(q) = \sum a_n q^n$ then we have

$$(Tr_{\mathrm{Frob}} f)(q) = p \sum a_{np} q^n \in pB[[q]].$$

By the q-expansion principle, it follows that "$\frac{1}{p} Tr_{\mathrm{Frob}} f$" is well defined, so:

Definition II.3.1 *Let $f \in \mathbf{V} = \mathbf{V}(B, \mathrm{N})$. We define $\mathrm{U}f \in \mathbf{V}$ to be the unique element of $\mathbf{V}$ satisfying*

$$p \cdot (\mathrm{U}f)(q) = (Tr_{\mathrm{Frob}} f)(q).$$

This defines a linear operator $\mathrm{U} : \mathbf{V} \longrightarrow \mathbf{V}$, which acts on q-expansions by

$$\sum a_n q^n \xrightarrow{\mathrm{U}} \sum a_{np} q^n,$$

and satisfies the relation $\mathrm{U}(\mathrm{Frob}(f) \cdot g) = f\mathrm{U}(g)$ (check on q-expansions). Following Monsky in [Mons71], we call operators with this property *Dwork operators*[1]. In particular, setting $g = 1$, we have $\mathrm{U}(\mathrm{Frob} f) = f$ for any $f \in \mathbf{V}$; we will later explore the consequences of this property.

Remark: Since the operator U we have just defined coincides with the classical one on q-expansions, the two also coincide on classical forms, and in particular on divided congruences. By continuity, it follows that the U operator defined above coincides with the one obtained by an inverse limit procedure in Section II.1.2.

II.3.2 U and overconvergence

Since the U operator commutes with the diamond operators (because the Frobenius endomorphism does), U must preserve weights, and therefore maps the space of p-adic modular forms of weight k and growth condition $r = 1$ to itself (since this is just the space of generalized p-adic modular functions of weight k). More generally, U preserves the spaces of modular forms of weight χ (and growth condition $r = 1$). It is not,

[1] The relevance of this property is not completely clear. In [Mons71], Monsky showed that any Dwork operator on a "weakly complete, weakly finitely generated" space was automatically a "nuclear operator", i.e., had a spectral theory. The result is proved by constructing subspaces that are somehow analogous to the spaces of overconvergent forms (i.e., defined by requiring that power-series coefficients converge to zero "better than linearly"). Though Monsky's result *cannot* be applied directly to our situation, his proof served as a guide at many points.

however, at all clear that U preserves the space of overconvergent forms (and in fact it is false without qualification). The goal of this section is to try to understand the action of U on the various spaces of overconvergent forms. The difficulties are the same as for Frob: first, the question of the existence of the fundamental subgroup, and second (for the case where the weight is not zero), the problem of pulling back a non-vanishing differential via an isogeny of degree p.

We want to determine to what extent U preserves overconvergence; since Frob only had good properties with respect to overconvergence up to tensoring with the fraction field K of B, we expect the same sort of behavior in the current case. The first crucial result is due to Katz.

Proposition II.3.2 *Suppose* $\mathrm{N} \geq 3$ *and* $p \not| \mathrm{N}$. *Then, for any* $r \in B$ *such that* $\mathrm{ord}(r) < 1/(p+1)$, *the homomorphism*

$$\mathrm{Frob} : \mathsf{M}(B, 0, \mathrm{N}; r^p) \otimes K \longrightarrow \mathsf{M}(B, 0, \mathrm{N}; r) \otimes K$$

is finite and étale of rank p.

Proof: This is [Ka73, Theorem 3.10.1]. For $r = 1$, it is an immediate consequence of Proposition II.2.9 above. For $\mathrm{ord}(r) > 0$, the proof is more difficult, and involves interpreting the K-algebras in question as the coordinate algebras of the (affinoid) rigid analytic spaces classifying elliptic curves over K satisfying $1 \geq |\mathrm{E}_{p-1}| \geq |r^p|$ and $1 \geq |\mathrm{E}_{p-1}| \geq |r|$, respectively, and doing a delicate study of the kernel of multiplication by p in the formal group of a "not too supersingular" curve (we quote one of the results of this analysis, which is due to Lubin, in Theorem II.3.5 below). □

Given this result, we see that we can define Tr_{Frob} as the trace map on global sections of $\underline{\omega}^{\otimes k}$ defined by the finite étale map Frob. Thus,

$$Tr_{\mathrm{Frob}}(\mathsf{M}(B, k, \mathrm{N}; r) \otimes K) \subset \mathsf{M}(B, k, \mathrm{N}; r^p) \otimes K,$$

so that we have:

Corollary II.3.3 *Suppose* $\mathrm{N} \geq 3$ *and* $p \not| \mathrm{N}$. *Then, for any integer* k *and any* $r \in B$ *such that* $\mathrm{ord}(r) < 1/(p+1)$, *we have*

$$\mathrm{U}(\mathsf{M}(B, k, \mathrm{N}; r) \otimes K) \subset \mathsf{M}(B, k, \mathrm{N}; r^p) \otimes K.$$

We interpret this as saying that, up to tensoring with K, the operator U *improves* overconvergence. It is not immediately clear, however, that giving $\mathsf{M}(B, k, \mathrm{N}; r) \otimes K$ and $\mathsf{M}(B, k, \mathrm{N}; r^p) \otimes K$ their "natural" p-adic topologies (in which $\mathsf{M}(B, k, \mathrm{N}; r)$ and $\mathsf{M}(B, k, \mathrm{N}; r^p)$ are the closed unit balls) makes the linear map $\mathrm{U} : \mathsf{M}(B, k, \mathrm{N}; r) \otimes K \longrightarrow \mathsf{M}(B, k, \mathrm{N}, r^p) \otimes K$ a bounded map. (Recall that the q-expansion topology is strictly weaker than the "natural" topology, as we remarked above.) The case of weight zero has been dealt with by Katz:

Lemma II.3.4 *For any $r \in B$ with $\operatorname{ord}(r) < 1/(p+1)$, we have*

$$\mathrm{U}(\mathsf{M}(B,0,\mathrm{N};r)) \subset p^{-1}\mathsf{M}(B,0,\mathrm{N};r^p).$$

Proof: This is Lemma 3.11.4 of [Ka73], and of course a special case of the general result below. □

Thus, the linear map

$$\mathrm{U} : \mathsf{M}(B,0,\mathrm{N};r) \otimes K \longrightarrow \mathsf{M}(B,0,\mathrm{N};r^p) \otimes K$$

is bounded, and hence a continuous linear map of p-adic Banach spaces. The situation for weight $k \neq 0$ is more complicated, since then Frob is already not integral. We deal with this by looking a little more carefully at the Frobenius map in characteristic zero, or equivalently, on the moduli spaces of "not too supersingular" curves. The result we need is due to Lubin. Let Ω be the completion of the algebraic closure of K, and let B_∞ be its ring of integers. Let E/B_∞ be a supersingular elliptic curve over B_∞, and let T be a parameter for the formal group of E, normalized by $[\zeta](T) = \zeta T$ for every $(p-1)^{\mathrm{st}}$ root of unity ζ in $\mathbf{Z}_p$. Multiplication by p in the formal group is given by

$$[p](T) = pT + aT^p + \sum_{i=2}^{p} c_i T^{i(p-1)+1} + c_{p+1}T^{p^2} + \dots$$

with $\operatorname{ord}(a) > 0$ (because E is supersingular), $\operatorname{ord}(c_i) \geq 1$ for $i \not\equiv 1(\operatorname{mod} p)$, and $\operatorname{ord}(c_p) = 0$ (because the formal group is of height 2). Note that since $a \equiv \mathrm{E}_{p-1}(\mathrm{E},\omega) \pmod{\mathrm{p}}$ for any nonvanishing differential ω on E, we have that, if $\operatorname{ord}(a) < 1$, then $\operatorname{ord}(a) = \operatorname{ord}(\mathrm{E}_{p-1}(\mathrm{E},\omega))$. We want to determine the curves (if any) that are mapped to E by quotient by their fundamental group.

Theorem II.3.5 *Let $0 < \operatorname{ord}(a) < p/(1+p)$, so that the canonical subgroup $\mathrm{H}_0 \subset \mathrm{E}$ is defined, and let $\mathrm{H}_1, \mathrm{H}_2, \dots, \mathrm{H}_p$ be the other finite flat subgroup schemes of rank p of E. Then there exist precisely p curves $\mathrm{E}^{(i)}$ having $\operatorname{ord}(a^{(i)}) < 1/(1+p)$ such that*

$$\mathrm{E} = \mathrm{E}^{(i)}/\mathrm{H}_0(\mathrm{E}^{(i)}),$$

where $\mathrm{H}_0(\mathrm{E}^{(i)})$ denotes the fundamental subgroup of $\mathrm{E}^{(i)}$. These are precisely the curves

$$\mathrm{E}^{(i)} = \mathrm{E}/\mathrm{H}_i,$$

$i = 1, 2, \dots, p$. Furthermore, we have

$$\operatorname{ord}(a^{(i)}) = \frac{1}{p}\operatorname{ord}(a).$$

Proof: This follows from a careful analysis of the kernel of multiplication by p in the formal group of E. See [Ka73, Thm. 3.10.7], where several other cases are examined, shedding some light on the question of the existence of the fundamental group. □

Since it is sufficient to prove U is bounded after tensoring with a large extension of B (e.g., because the conditions in Corollary I.2.8 are independent of the ring B), we may work over B_∞, and apply the previous result to deal with the question of which curves are mapped to a given curve by quotient by their fundamental group. To deal with the problem of pulling back the differential, we use the theory of the Hasse invariant and the fact that it is congruent modulo p to E_{p-1}.

Proposition II.3.6 *Let* $\mathrm{N} \geq 3$, $p \not| \mathrm{N}$, *and assume that either* $k \neq 1$ *or* $\mathrm{N} \leq 11$. *Let* $f \in \mathsf{M}(B_\infty, k, \mathrm{N}; r)$, *for any* $r \in B$ *such that* $\mathrm{ord}(r) < 1/(p+1)$. *Then* $Tr_{\mathrm{Frob}} f \in \mathsf{M}(B_\infty, k, \mathrm{N}; r^p)$.

Proof: We want to define the value of $Tr_{\mathrm{Frob}} f$ on any test object $(\mathrm{E}/B_\infty, \omega, \imath, Y)$ of level N and growth condition r^p; our strategy is to compute formally and hope for the best. Thus, formally, we would have

$$(Tr_{\mathrm{Frob}} f)(\mathrm{E}, \omega, \imath, Y) = \sum f(\mathrm{E}_1, \omega_1, \imath_1, Y_1),$$

where the sum is over the $(\mathrm{E}_1, \omega_1, \imath_1, Y_1)$ which map to $(\mathrm{E}, \omega, \imath, Y)$ by division by the fundamental subgroup. The point of the proof is to show that we can give a sense to the expression inside the summation above. We do this by an argument similar to that in the proof of Theorem 3.3 in [Ka73, p.46].

As a first step, we determine explicitly the triples

$$(\mathrm{E}_1, \omega_1, \imath_1, Y_1)$$

lying over the given triple $(\mathrm{E}, \omega, \imath, Y)$. It is clear from the Theorem II.3.5 that the only possibilities are triples of the form $(\mathrm{E}_i, \omega_1, \check{\pi} \circ \imath, Y_1)$, where $\mathrm{E}_i = \mathrm{E}/\mathrm{H}_i$, and ω_1 and Y_1 must be correctly chosen. Let $\pi_i : \mathrm{E}_i \longrightarrow \mathrm{E}$ be the projection on the quotient by the fundamental group. The main difficulty is that, since neither π_i nor its dual are étale when E is supersingular, we do not know if the pullback of a non-vanishing differential is still non-vanishing (and it will usually not be so). In any case, we may choose $\lambda_i \in B_\infty$ so that $\pi_i^*\omega = \lambda_i\omega_i$, where ω_i is a non-vanishing differential on E_i. Note then that if we write $\check{\pi}_i^*\omega_i = \lambda\omega$ we have $\lambda_i\lambda = p$.

Since $\mathrm{ord}(r) < 1/(p+1)$, $r_1 = p/r$ has $\mathrm{ord}(r_1) > p/(p+1)$ and hence is divisible by r^p; set $r_2 = r_1/r^p$, and note that $\mathrm{ord}(r_2) > 0$. By Theorem II.2.2, E mod $r_1 B$ is the Frobenius transform $(\mathrm{E}_i)^{(p)}$ of E_i, so that we may choose λ_i above so that ω reduces modulo r_1 to $\omega_i^{(p)}$ on $(\mathrm{E}_i)^{(p)}$. Then, as Katz shows in [Ka73, p.54], if $Y' \cdot \mathrm{E}_{p-1}(\mathrm{E}_i, \omega_i)$, we must have

$$(Y')^p = \frac{Y}{1 - r_2 Y}.$$

Since we are working over B_∞, we can solve for Y' by choosing the unique solution of the above equation satisfying $Y' \cdot \mathrm{E}_{p-1}(\mathrm{E}_i, \omega_i) = r$.

Since we have $\check{\pi}^*\pi^*\omega = p\omega$, we must have, formally,

$$\omega_1 = \frac{1}{p}\pi^*\omega = \frac{1}{p}\lambda_i\omega_i = \frac{1}{\lambda}\omega_i,$$

and

$$Y_1 = \frac{1}{\lambda^{p-1}} Y'.$$

Hence, formally, we have

$$\begin{aligned} f(\mathrm{E}_1, \omega_1, \imath_1, Y_1) &= f(\mathrm{E}_i, \frac{1}{\lambda}\omega_i, \check{\pi} \circ \imath, \frac{1}{\lambda^{p-1}} Y') \\ &= \lambda^k f(\mathrm{E}_i, \omega_i, \check{\pi} \circ \imath, Y') \\ &= \left(\frac{\lambda}{\mathrm{E}_{p-1}(\mathrm{E}_i, \omega_i)} \right)^k (\mathrm{E}_{p-1}^k f)(\mathrm{E}_i, \omega_i, \check{\pi} \circ \imath, Y'). \end{aligned}$$

This is of course only a formal computation because λ is *not* a unit in B_∞; however, the last term *is* well-defined and independent of the choices made because E_{p-1} is a lifting of the Hasse invariant, hence differs from λ by multiplication by a unit of B, (since $\mathrm{ord}(\mathrm{E}_{p-1}(\mathrm{E}_i, \omega_i)) < 1$, all liftings of the Hasse invariant have the same valuation), and this unit can be interpreted without ambiguity by "reduction to the universal case", in which B is flat over $\mathbf{Z}_p$. One can then check without difficulty that this transforms as expected when we multiply ω by a unit in B_∞. Thus, we obtain a well-defined value for $Tr_{\mathrm{Frob}} f$, which coincides with the previous one when the curve is not supersingular (because then all of our computations make sense!). This proves the proposition. □

Hence we have:

Corollary II.3.7 *Under the hypotheses of Proposition II.3.6, the map*

$$\mathrm{U} : \mathsf{M}(B, k, \mathrm{N}; r) \otimes K \longrightarrow \mathsf{M}(B, k, \mathrm{N}, r^p) \otimes K$$

is a bounded homomorphism of p-adic Banach spaces.

For the rest of this section, we assume that $k \neq 1$ or that $\mathrm{N} \leq 11$, so that we are in the situation where Proposition II.3.6 holds. Then, if $\mathrm{ord}(r) < 1/(p+1)$, the U operator induces a bounded linear map

$$\mathsf{M}(B, k, \mathrm{N}; r) \otimes K \longrightarrow \mathsf{M}(B, k, \mathrm{N}; r^p) \otimes K.$$

In fact, we know that

$$\mathrm{U}(\mathsf{M}(B, k, \mathrm{N}; r)) \subset \frac{1}{p} \mathsf{M}(B, k, \mathrm{N}; r^p),$$

so that $\|\mathrm{U}\| \leq p$. Finally, since we have $\mathsf{M}(B, k, \mathrm{N}; r^p) \hookrightarrow \mathsf{M}(B, k, \mathrm{N}; r)$, it is clear that we can consider U as a continuous linear endomorphism of the Banach space $\mathsf{M}(B, k, \mathrm{N}; r) \otimes K$. In fact, we can do better than this, by considering the composition

$$\mathsf{M}(B, k, \mathrm{N}; r^p) \otimes K \hookrightarrow \mathsf{M}(B, k, \mathrm{N}; r) \otimes K \xrightarrow{\ \mathrm{U}\ } \mathsf{M}(B, k, \mathrm{N}; r^p) \otimes K,$$

which shows that U defines a continuous linear endomorphism of the Banach space $\mathsf{M}(B, k, \mathrm{N}; r^p) \otimes K$. Finally, given any $r_1 \in B$ with $\mathrm{ord}(r_1) < p/(p+1)$, we may assume (since overconvergence properties can be detected by the congruence conditions of Corollary I.2.8, which are independent of the base ring) that $r_1 = r^p$ for some $r \in B$, and apply the preceding observation. Hence we have:

Corollary II.3.8 *Under the above hypotheses, and if $r \in B$ with $\operatorname{ord}(r) < p/(p+1)$, the* U *operator induces a continuous linear endomorphism of the p-adic Banach space* $\mathsf{M}(B, k, \mathrm{N}; r) \otimes K$.

We have seen that U is integral (i.e., $\|\mathrm{U}\| \leq 1$) on $\mathbf{V}$ (and hence on the spaces of forms with growth condition $r = 1$). This is *not* true in general. However, for r in the appropriate range, it is almost true, in the sense that there there is an equivalent p-adic metric on the space $\mathsf{M}(B, k, \mathrm{N}; r) \otimes K$ for which U is of norm one. Let

$$L_k(r) = \mathsf{M}(B, k, \mathrm{N}; r) + U(\mathsf{M}(B, k, \mathrm{N}; r)),$$

Then, since

$$\mathrm{U}(\mathsf{M}(B, k, \mathrm{N}; r)) \subset \frac{1}{p}\mathsf{M}(B, k, \mathrm{N}; r),$$

we have

$$\mathsf{M}(B, k, \mathrm{N}; r) \subset L_k(r) \subset \frac{1}{p}\mathsf{M}(B, k, \mathrm{N}; r),$$

so that taking $L_k(r)$ as the unit ball defines a p-adic topology on the Banach space $\mathsf{M}(B, k, \mathrm{N}; r) \otimes K$ which is *equivalent* to the canonical one. Then we have the following result, which generalizes a result of Dwork to the case of weight $k \neq 0$ (the case of weight zero is [Ka73, Lemma 3.11.7]):

Proposition II.3.9 *With hypotheses and definitions as above, assume that $p \geq 7$ and*

$$\frac{2p}{3(p-1)} < \operatorname{ord}(r) < \frac{p}{p+1}.$$

Then we have $\mathrm{U}(L_k(r)) \subset L_k(r)$, so that, in the p-adic norm defined by $L_k(r)$ we have $\|\mathrm{U}\| \leq 1$.

Proof: As pointed out above, we may as well assume that $r = r_1^p$, for r_1 satisfying

$$\frac{2}{3(p-1)} < \operatorname{ord}(r_1) < \frac{1}{p+1}.$$

(The hypothesis $p \geq 7$ is needed for this inequality to be possible.)

It is sufficient to prove that $f \in \mathsf{M}(B, k, \mathrm{N}; r)$ implies $\mathrm{U}^2(f) \in L_k(r)$. The proof, which is essentially the same as that given by Katz, amounts to using Corollary I.2.8 repeatedly to determine overconvergence. Thus, let $f \in \mathsf{M}(B, k, \mathrm{N}; r)$; by Proposition I.2.6, we may write

$$f = b_0 + \frac{rb_1}{\mathrm{E}_{p-1}} + \frac{r^2 b_2}{\mathrm{E}_{p-1}^2} + \cdots$$

Since $b_0 \in M(B, k, \mathrm{N})$ is classical,

$$\mathrm{U}(b_0) \in M(B, k, \Gamma_1(\mathrm{N}) \cap \Gamma_0(p)) \subset \mathsf{M}(B, k, \mathrm{N}; r),$$

so that $\mathrm{U}^2(b_0) \in L_k(r)$, and we need not worry about the first term. Next, note that

$$\operatorname{ord}(\frac{r^i}{pr_1^i}) = \operatorname{ord}(\frac{r_1^{pi}}{pr_1^i}) = i(p-1)\operatorname{ord}(r_1) - 1 > 0,$$

for any $i \geq 2$, so that we may write

$$f = b_0 + \frac{rb_1}{\mathrm{E}_{p-1}} + p \cdot g,$$

with $g \in \mathfrak{m}\mathsf{M}(B,k,\mathrm{N};r_1)$, where $\mathfrak{m}$ denotes the maximal ideal of B (just factor out p), so that $\mathrm{U}(p\cdot g) = p\mathrm{U}(g) \in \mathsf{M}(B,k,\mathrm{N};r)$ and hence $\mathrm{U}^2(p\cdot g) \in L_k(r)$. Hence, it remains to show that

$$\mathrm{U}^2(\frac{rb_1}{\mathrm{E}_{p-i}}) \in L_k(r).$$

Writing

$$\frac{rb_1}{\mathrm{E}_{p-1}} = \frac{r_1^p b_1}{\mathrm{E}_{p-1}} = r_1^{p-1}\frac{r_1 b_1}{\mathrm{E}_{p-1}},$$

and noting that

$$\mathrm{U}(\frac{r_1 b_1}{\mathrm{E}_{p-1}}) \in \frac{1}{p}\mathsf{M}(B,k,\mathrm{N},r_1^p),$$

we can write

$$\begin{aligned}
\mathrm{U}(\frac{rb_1}{\mathrm{E}_{p-1}}) &= r_1^{p-1}\mathrm{U}(\frac{r_1 b_1}{\mathrm{E}_{p-1}}) \\
&= \frac{r_1^{p-1}}{p}\left(b_0' + \frac{r_1^p b_1'}{\mathrm{E}_{p-1}} + \frac{r_1^{2p} b_2'}{\mathrm{E}_{p-1}^2} + \cdots\right) \\
&= \frac{r_1^{p-1}}{p}\left(b_0' + \frac{r_1^p b_1'}{\mathrm{E}_{p-1}}\right) + p\cdot h,
\end{aligned}$$

where, since

$$\operatorname{ord}(\frac{r_1^{ip+p-1}}{p^2 r_1^i}) = (i+1)(p-1)\operatorname{ord}(r_1) - 2 > 0$$

for $i \geq 2$, we have $h \in \mathfrak{m}\mathsf{M}(B,k,\mathrm{N};r_1)$. Then, as before, $\mathrm{U}(p\cdot h) \in L_k(r)$, and it remains to show that

$$\mathrm{U}\left(b_0' + \frac{r_1^p b_1'}{\mathrm{E}_{p-1}}\right) \in L_k(r).$$

Now, the q-expansion of $\mathrm{U}(\frac{r_1^p b_1}{\mathrm{E}_{p-1}})$ is divisible by $r = r_1^p$ (because U is integral on $\mathbf{V}$, and hence on q-expansions), and the q-expansion of $p\cdot h$ is divisible by p, and hence also divisible by r. Hence, r must divide the q-expansion of

$$\frac{r_1^{p-1}}{p}\left(b_0' + \frac{r_1^p b_1'}{\mathrm{E}_{p-1}}\right) = \frac{r_1^{p-1}}{p}\left(\frac{b_0'\mathrm{E}_{p-1} + r_1^p b_1'}{\mathrm{E}_{p-1}}\right),$$

and therefore also the q-expansion of

$$\frac{r_1^{p-1}}{p}(b_0'\mathrm{E}_{p-1} + r_1^p b_1'),$$

which is *classical* of weight $k+p-1$. By the q-expansion principle for classical forms, there exists a classical modular form b_1'' of weight $k+p-1$ (and level N) such that

$$\frac{r_1^{p-1}}{p}(b_0'\mathrm{E}_{p-1} + r_1^p b_1') = r_1^p b_1'' = r b_1'',$$

and so

$$\frac{r_1^{p-1}}{p}\left(b_0' + \frac{r_1^p b_1'}{\mathrm{E}_{p-1}}\right) = \frac{r b_1''}{\mathrm{E}_{p-1}} \in \mathsf{M}(B,k,\mathrm{N};r).$$

Then clearly

$$\mathrm{U}\left(\frac{r_1^{p-1}}{p}(b_0' + \frac{r_1^p b_1'}{\mathrm{E}_{p-1}})\right) = \mathrm{U}\left(\frac{r b_1''}{\mathrm{E}_{p-1}}\right) \in L_k(r),$$

proving our claim. □

A closer look at the proof will reveal that we have also proved two congruence properties, namely:

Corollary II.3.10 *Under the above hypotheses, write*

$$f = \sum \frac{r^a b_a}{\mathrm{E}_{p-1}^a},$$

and let $\mathfrak{m}$ *denote the maximal ideal of* B. *Then we have*

$$\mathrm{U}(f) \equiv \mathrm{U}(b_0) + \mathrm{U}(\frac{r b_1}{\mathrm{E}_{p-1}}) \pmod{\mathfrak{m}},$$

and, for some $b_1'' \in M(B, k+p-1, \mathrm{N})$,

$$\mathrm{U}^2(f) \equiv \mathrm{U}(\frac{r b_1''}{\mathrm{E}_{p-1}}) \pmod{\mathfrak{m}}.$$

Proof: The first statement is clear from the proof of Proposition II.3.9, as is the fact that

$$\mathrm{U}^2(f) \equiv \mathrm{U}^2(b_0) + \mathrm{U}(\frac{r b_1'}{\mathrm{E}_{p-1}}) \pmod{\mathfrak{m}},$$

for some $b_1' \in M(B, k+p-1, \mathrm{N})$. Then, noting that $\mathrm{U}(b_0) \in \mathsf{M}(B,k,\mathrm{N};r)$ (because b_0 is classical), and applying the first statement to $\mathrm{U}(b_0)$ yields the second statement at once. □

Note that these results show that $L_k(r)$ is a U-stable B-lattice

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

and that the action of U on $L_k(r) \otimes B/\mathfrak{m}$ is determined by its action on the classical space

$$M(B, k+p-1, \mathrm{N}) \otimes B/\mathfrak{m};$$

for example, the unit eigenvalues of U acting on $L_k(r)$ will be congruent to unit eigenvalues of U acting on the space of classical modular forms of weight $k+p-1$ and level N (in fact, Hida has shown in [Hi86b] that for $k \geq 3$ any $f \in \mathsf{M}(B, k, \mathrm{N}; 1)$ having a unit eigenvalue under U is necessarily a *classical* modular form of weight k and level N).

It is interesting that one can prove integrality results for the U operator for $\mathrm{ord}(r) = 0$ (i.e., for $r = 1$) and for

$$\frac{2p}{3(p-1)} < \mathrm{ord}(r) < \frac{p}{p+1},$$

which are, so to speak, the "opposite ends" of the range of r for which U gives an endomorphism of $\mathsf{M}(B, k, \mathrm{N}; r)$. It seems that some integrality result must be possible in all cases in that range; we will later look at another part of the range to estimate the Newton polygon of the characteristic power series of U.

II.3.3 U and Frobenius

The point of this section is to exploit to the fullest the identity $\mathrm{U}(\mathrm{Frob}f) = f$ and its variants. The most important consequence, from the point of view of the rest of this chapter, will be to show that the spectral theory of the U operator is dramatically different in the overconvergent and the non-overconvergent cases. Specifically, we will see in the next section that the U operator is completely continuous on the p-adic Banach spaces of overconvergent modular forms, so that its eigenvalues form a sequence tending to zero. By contrast, we show in this section that *every* element of the maximal ideal of the completion of the ring of integers of a separable closure of $\mathbf{Q}_p$ is an eigenvalue for the U operator acting on the full space of p-adic modular forms of weight k.

We begin with a simple consequence of the expressions of U and Frob on q-expansions:

Proposition II.3.11 *Let B be a p-adic ring, and* N *an integer prime to p. The sequence*

$$0 \longrightarrow \mathbf{V}(B, \mathrm{N}) \xrightarrow{\mathrm{Frob}} \mathbf{V}(B, \mathrm{N}) \xrightarrow{1-\mathrm{Frob}\circ\mathrm{U}} \mathbf{V}(B, \mathrm{N}) \xrightarrow{\mathrm{U}} \mathbf{V}(B, \mathrm{N}) \longrightarrow 0$$

is exact.

Proof: We check exactness at each step:

a) Frob is injective, since, for $f(q) = \sum a_n q^n$, $(\mathrm{Frob}(f))(q) = \sum a_n q^{np} = 0$ implies $a_n = 0$ for all n, and hence $f = 0$ by the q-expansion principle.

b) First, $(1 - \mathrm{Frob}\circ\mathrm{U})(\mathrm{Frob}(f)) = \mathrm{Frob}(f) - \mathrm{Frob}(\mathrm{U}(\mathrm{Frob}(f))) = \mathrm{Frob}(f) - \mathrm{Frob}(f) = 0$. Conversely, if $(1 - \mathrm{Frob}\circ\mathrm{U})(f) = 0$, then $f = \mathrm{Frob}(\mathrm{U}(f))$.

c) First, $\mathrm{U}(f - \mathrm{Frob}(\mathrm{U}(f))) = \mathrm{U}(f) - \mathrm{U}(f) = 0$. Conversely, if $\mathrm{U}(f) = 0$, then

$\mathrm{Frob}(\mathrm{U}(f)) = 0$, so that $f = (1 - \mathrm{Frob}\circ\mathrm{U})(f)$.

d) Finally U is surjective because $f = \mathrm{U}(\mathrm{Frob}(f))$. □

Of course, since, with the usual hypotheses on B and N, both U and Frob preserve the space of modular forms of weight k, we also have:

Corollary II.3.12 *Let B be a p-adically complete discrete valuation ring, let $\mathrm{N} \geq 3$, $p \nmid \mathrm{N}$, and assume that either $k \neq 1$ or $\mathrm{N} \leq 11$. Then we have an exact sequence:*

$$0 \longrightarrow \mathsf{M}(B,k,\mathrm{N};1) \xrightarrow{\mathrm{Frob}} \mathsf{M}(B,k,\mathrm{N};1) \xrightarrow{1-\mathrm{Frob}\circ\mathrm{U}} \mathsf{M}(B,k,\mathrm{N};1) \xrightarrow{\mathrm{U}} \mathsf{M}(B,k,\mathrm{N};1) \longrightarrow 0.$$

A version with the spaces $\mathsf{M}(B,k,\mathrm{N};r) \otimes K$ of overconvergent modular forms of weight k also follows, with the obvious caution as to the degree of overconvergence at each step, as in the previous section. For this section, we concentrate on the larger space, and we fix the above hypotheses on B, N and k.

We should note in particular that the operator $(1 - \mathrm{Frob}\circ\mathrm{U})$ is in fact idempotent. Thus, we get:

Corollary II.3.13 *With the hypotheses above, we have a direct sum decomposition:*

$$\mathsf{M}(B,k,\mathrm{N};1) = \mathrm{image}(\mathrm{Frob}) \oplus \ker(\mathrm{U}).$$

Furthermore,

$$\ker \mathrm{U} = \mathrm{image}(1 - \mathrm{Frob}\circ\mathrm{U}) \cong \mathsf{M}(B,k,\mathrm{N};1) \big/ \mathrm{image}(\mathrm{Frob}),$$

so that the kernel of U *is infinite-dimensional.*

Since the operator $(1 - \mathrm{Frob}\circ\mathrm{U})$ maps the space $\mathsf{M}(B,k,\mathrm{N};r) \otimes K$ to itself (provided $\mathrm{ord}(r) < p/(p+1)$), the analogous result also holds in the overconvergent case.

To sum up, given any p-adic modular form $f \in \mathsf{M}(B,k,\mathrm{N};1)$, one can produce an eigenform for the U operator (with eigenvalue 0) simply by taking $f_0 = (1 - \mathrm{Frob}\circ\mathrm{U})(f)$. One should note, also, that the q-expansions of f and f_0 agree "outside p", i.e., if $a_n(g)$ denotes the coefficient of q^n in the q-expansion of g, we have $a_n(f) = a_n(f_0)$ whenever $p \nmid n$. In particular, if f is an eigenform for the Hecke operators T_ℓ for $\ell \neq p$, then so is f_0, and the eigenvalues are the same.

In fact, one can go much further than this, by the following construction. Let $f_0 \in \mathsf{M}(B,k,\mathrm{N};1)$ be such that $\mathrm{U}(f_0) = 0$. Take any $\lambda \in B$ such that $\mathrm{ord}(\lambda) > 0$, and consider the p-adic modular form

$$f_\lambda = f_0 + \lambda\mathrm{Frob}(f_0) + \lambda^2\mathrm{Frob}^2(f_0) + \ldots + \lambda^n\mathrm{Frob}^n(f_0) + \ldots$$

Note, first, that since $\lambda^n \to 0$, the series clearly converges and defines an element of $\mathsf{M}(B,k,\mathrm{N};1)$. Furthermore,

$$\begin{aligned} \mathrm{U}f_\lambda &= \mathrm{U}f_0 + \lambda\mathrm{U}(\mathrm{Frob}f_0) + \lambda^2\mathrm{U}(\mathrm{Frob}^2 f_0) + \ldots \\ &= 0 + \lambda f_0 + \lambda^2\mathrm{Frob}f_0 + \ldots \\ &= \lambda f_\lambda. \end{aligned}$$

Thus, given *any* $f_0 \in \ker(\mathrm{U})$ and *any* λ in the maximal ideal of B, we have constructed a p-adic modular form $f_\lambda \in \ker(\mathrm{U}-\lambda)$; furthermore, f_0 and f_λ clearly have the same q-expansion coefficients "outside of p". Thus, we have proved:

Proposition II.3.14 *With the hypotheses above, for any λ in the maximal ideal of B, there is a bicontinuous bijection*

$$\begin{array}{ccc} \ker(\mathrm{U}) & \longrightarrow & \ker(\mathrm{U}-\lambda) \\ f_0 & \mapsto & f_\lambda . \end{array}$$

Furthermore, we also have $a_n(f_0) = a_n(f_\lambda)$ whenever $p \nmid n$. Thus, if f_0 is an eigenform for the Hecke operators T_ℓ with $\ell \neq p$, then so is f_λ, and with the same eigenvalues.

Remarks:

1) If we look at the above construction from the point of view of q-expansions, it may be interpreted in very naïve terms. Suppose we wish to construct the q-expansion of an eigenform for the U operator with eigenvalue λ. Then we might proceed as follows:

i. fix the a_n with $p \nmid n$ arbitrarily (this corresponds to choosing an f_0);

ii. when $p|n$ but $p^2 \nmid n$, set $a_n = \lambda a_{n/p}$;

iii. when $p^2|n$ but $p^3 \nmid n$, set $a_n = \lambda^2 a_{n/p^2} = \lambda a_{n/p}$;

iv. in general, if $n = p^\nu m$ and $p \nmid m$, set $a_n = \lambda^\nu a_m$.

Notice that this makes perfect sense for *any* λ in B. The point of the above discussion, then, is that if we begin with a p-adic modular form $f_0 \in \ker(\mathrm{U})$ and if λ is in the maximal ideal of B, then the resulting q-expansion is in fact the q-expansion of a p-adic modular form. If λ is a unit in B, this is *not* necessarily the case, as we shall see shortly. (Specifically, what we shall show is that for each fixed weight there are only a finite number of pairs (f_0, λ) for which λ is a unit in B and the q-expansion we have just constructed is the q-expansion of a p-adic modular form.)

2) The fact that the Frobenius endomorphism maps the space $\mathsf{M}(B, k, \mathrm{N}; r^p) \otimes K$ to $\mathsf{M}(B, k, \mathrm{N}; r) \otimes K$ (i.e., it reduces overconvergence) shows that even if we start with an overconvergent p-adic modular form f_0, we cannot guarantee that f_λ will be also overconvergent. In fact, what is true, as will follow from the results in the next section, is that for each fixed weight there is only a denumerable set of pairs (f_0, λ) for which $\lambda \neq 0$ and the above construction produces the q-expansion of an overconvergent modular form, and the possible values of λ form a sequence tending to zero. (It would be quite interesting to obtain an a priori criterion for determining whether a pair (f_0, λ) is of this kind, if one exists!)

To summarize, one might say that to consider the spectral theory of the U operator on the full space $\mathsf{M}(B,k,\mathrm{N};1)$ produces very little information, except for the ordinary part, i.e., the spectral theory for eigenvalues which are units in B. In the next section, we consider the spectral theory of U acting on the space of overconvergent forms, and show that this is in fact much more interesting.

II.3.4 Spectral theory: the overconvergent case

In this section we study the spectral theory of the U operator acting on overconvergent forms of integral weight k. The fundamental result is that the U operator acting on spaces of overconvergent p-adic modular forms is a completely continuous operator (so that there *is* a spectral theory to study). This turns out to follow immediately from the fact that U improves overconvergence.

Proposition II.3.15 *Let* $\mathrm{N} \geq 3$, $p \nmid \mathrm{N}$, *and assume that either* $k \neq 1$ *or* $\mathrm{N} \leq 11$. *Let* B *be a* p*-adically complete discrete valuation ring such that* B/pB *is finite, and let* $r \in B$ *satisfy* $0 < \operatorname{ord}(r) < p/(p+1)$. *Then the operator*

$$\mathrm{U} : \mathsf{M}(B,k,\mathrm{N};r) \otimes K \longrightarrow \mathsf{M}(B,k,\mathrm{N};r) \otimes K$$

is completely continuous.

Proof: As before, write $r = r_1^p$. Then U, considered as an endomorphism of $\mathsf{M}(B,k,\mathrm{N};r) \otimes K$, factors as

$$\mathsf{M}(B,k,\mathrm{N};r) \otimes K \hookrightarrow \mathsf{M}(B,k,\mathrm{N};r_1) \otimes K \xrightarrow{\ \mathrm{U}\ } \mathsf{M}(B,k,\mathrm{N};r) \otimes K.$$

By Corollary I.2.9 the inclusion is a completely continuous homomorphism of p-adic Banach spaces, and the corollary follows. □

It follows that the U operator has all the properties of completely continuous operators on p-adic Banach spaces; since many of these are crucial to our theory, we will discuss them in more detail. The reference for all of our statements is the paper [Se62] of Serre (see also the remarks in Monsky's paper [Mons71]). The first important result is that there is a spectral theory completely analogous to the classical one. Let $g(X) \in K[X]$ be a polynomial with $g(0) \neq 0$; then we have

$$\mathsf{M}(B,k,\mathrm{N};r) \otimes K = M(g) \oplus F(g),$$

where $g(\mathrm{U})$ is bijective and bicontinuous on $F(g)$ and $g(\mathrm{U})^n$ annihilates $M(g)$ for some n. Of course, the most common example is $g(X) = X - \lambda$, where λ is an eigenvalue of U, in which case $M(g) = M(\lambda)$ is the generalized eigenspace corresponding to the eigenvalue λ. The point of extending this to polynomials is that it allows us to consider the case when the eigenvalues do not belong to the field K, by taking $g(X)$ to be the minimal polynomial.

In fact, we can even extend this to sets of polynomials (essentially by using the fact that $K[X]$ is noetherian): let $S \subset K[X]$ be a set of polynomials, and assume that the roots in $\overline{K}$ of the polynomials in S are bounded away from 0 (so that in particular S is disjoint from the ideal $X \cdot K[X]$); then $M(S) = \sum_{g\in S} M(g)$ is finite-dimensional over K, and we have a direct sum decomposition

$$\mathsf{M}(B, k, \mathrm{N}; r) \otimes K = M(S) \oplus F(S).$$

Using the existence of the spectral decomposition, we can then define the trace and the characteristic power series of the operator U. For any S as above, define $Tr_S(\mathrm{U}^n) = \mathrm{trace}(\mathrm{U}^n|M(S))$ and $P_S(t) = \det(1 - t\mathrm{U}|M(S))$. Noting that the family of such S form a directed set under inclusion, we can view the maps $S \to Tr_S(\mathrm{U}^n)$ and $S \to P_S(t)$ as nets in K and in $K[[t]]$, respectively, where we give $K[[t]]$ the topology of coefficientwise convergence. Then the limits $Tr(\mathrm{U}^n) = \lim_S Tr_S(\mathrm{U}^n)$ and $P(t) = \lim_S P_S(t)$ both exist. The resulting power series $P(t)$ is the p-adic analogue the Fredholm determinant $\det(1 - t\mathrm{U})$ (in the sense of Serre in [Se62]; note, however, that the construction given above is different from Serre's, and goes through whenever there is a spectral theory; when the operator in question is completely continuous, it is equivalent to Serre's construction). Hence, in particular, $\lambda \neq 0$ is as eigenvalue of U if and only if $P(\lambda^{-1}) = 0$ and the dimension of the generalized eigenspace corresponding to λ is precisely the multiplicity of λ^{-1} as a root of $P(t)$.

The next important remark is that the power series $P(t)$ defines a p-adic entire function. To be specific, we have

$$P(t) = \exp(-\sum_{n=1}^{\infty}(\frac{Tr(\mathrm{U}^n)t^n}{n})) = 1 - Tr(\mathrm{U})t + \ldots = \sum c_i t^i,$$

with

$$\lim \frac{\mathrm{ord}(c_i)}{i} = \infty,$$

so that $P(t)$ is entire. Hence, we may write

$$P(t) = \prod_i (1 - \lambda_i t),$$

with $\lambda_i \to 0$ where $\lambda_i \in \overline{K}$ (the algebraic closure of K) are the eigenvalues of U. In the same way, we may define, following Serre, the Fredholm resolvent

$$F(t, \mathrm{U}) = \frac{P(t)}{1 - t\mathrm{U}} = \frac{\det(1 - t\mathrm{U})}{1 - t\mathrm{U}},$$

as a formal power series whose coefficients are polynomials in U, and which again is "entire", in the sense that, for every $\mu \in K$, the series $F(\mu, \mathrm{U})$ converges in the norm topology for operators. This allows us to show:

Lemma II.3.16 *Given g as in (1), let ψ_g be the inverse of $g(\mathrm{U})$ on the subspace $F(g)$, i.e., the function defined by $\psi_g|M(g) = 0$ and $\psi_g|F(g) = (g(\mathrm{U})|F(g))^{-1}$, let π_g be the "projection onto $F(g)$" (which is given by $\psi_g^n \circ g(\mathrm{U})^n$ for n sufficiently large), let $\mathcal{A}$ be the subalgebra of the algebra of continuous endomorphisms of $M(B,k,\mathrm{N};r)\otimes K$ generated by the identity and U, and let $\overline{\mathcal{A}}$ be its closure in the norm topology. Then $\psi_g \in \overline{\mathcal{A}}$ and $\pi_g \in \overline{\mathcal{A}}$.*

Proof: This is implicit in [Se62], as is noted by Monsky in [Mons71]. The statements for π_g and for ψ_g are equivalent, and we will concentrate on the projection π_g.

Consider first the case where $g(X) = X - \lambda$. Let h denote the multiplicity of λ^{-1} as a root of $P(t)$, and let Δ denote the operator on formal power series defined by

$$\Delta^s H(t) = \frac{1}{s!}\frac{d^s}{dt^s}H(t).$$

Then Serre shows that π_g is the h^{th} power of the operator given by

$$(\Delta^h P(\lambda^{-1}))^{-1}(1-\lambda^{-1}\mathrm{U})\Delta^h F(\lambda^{-1},\mathrm{U}),$$

where $F(t,\mathrm{U})$ denotes the Fredholm resolvent, as above. Since all the power series involved are entire, this gives a power series in U which converges in the norm topology (since U is essentially integral, this just means that the coefficients tend to zero!). This proves the assertion for the projection onto $F(g)$, when $g(X) = X - \lambda$. The general case then follows by writing $g(\mathrm{U}) = \lambda - \mathrm{U}_1$ for some $\lambda \in K$ and a completely continuous operator U_1. The assertion for ψ_g is proved similarly; we refer the reader to Serre's paper for the details. □

Corollary II.3.17 *For any g as above, the operators e_g and ψ_g commute with the Hecke operators T_ℓ on $\mathsf{M}(B,k,\mathrm{N};r)\otimes K$.*

Thus, we obtain:

Corollary II.3.18 *Let $p \geq 7$, $\mathrm{N} \geq 3$, and assume $k \neq 1$ or $\mathrm{N} \leq 11$. Let $g(X)$ be a polynomial with nonzero independent term, and let r be as in Proposition II.3.9.*

Then we have a decomposition

$$\mathsf{M}(B,k,\mathrm{N};r)\otimes K = M(g) \oplus F(g)$$

such that $g(\mathrm{U})$ is nilpotent on $M(g)$ and invertible with continuous inverse on $F(g)$. The space $M(g)$ is finite-dimensional, independent of r such that $\mathrm{ord}(r) < p/(p+1)$, and consists of the overconvergent modular forms of weight k on which $g(\mathrm{U})$ is nilpotent.

The characteristic power series $P(t)$ of the U operator and the spaces $M(S)$ are independent of r with $0 < \mathrm{ord}(r) < p/(p+1)$. Moreover, $P(t)$ has integral coefficients, i.e., $P(t) \in B[[t]]$, so that the eigenvalues of U are all integral. Finally, for each $\alpha \geq 0$, the set of eigenvalues λ satisfying $0 \leq \mathrm{ord}(\lambda) \leq \alpha$ is finite.

Proof: The existence of the decomposition is, of course, an immediate consequence of the fact that U is completely continuous. To show that $M(g)$ is independent of r for any g as above, note that, since $g(0) \neq 0$, $g(U)^n(f) = 0$, $f \in M(B, k, \mathrm{N}; r) \otimes K$ implies that f belongs to the span of $\mathrm{U}f$, $\mathrm{U}^2 f$, etc., and hence that $f \in M(B, k, \mathrm{N}; r^p) \otimes K$; by the same argument we get $f \in M(B, k, \mathrm{N}; r^{p^2}) \otimes K$, and so on until we have $1/(p+1) < \mathrm{ord}(r^{p^n}) < p/(p+1)$. It follows immediately that the same is true for any of the spaces $M(S)$, and hence that it is true for the characteristic power series $P(t)$.

That $P(t) \in B[[t]]$ now follows immediately by choosing the appropriate r and applying Proposition II.3.9. Finally, the last statement is a standard property of p-adic entire functions. □

It is useful to note that we may compute the characteristic power series $P(t) \in B[[t]]$ of U on any space $\mathsf{M}(B, k, \mathrm{N}; r) \otimes K$ for any r satisfying $0 < \mathrm{ord}(r) < p/(p+1)$, and call it *the* characteristic power series of U, since it is independent of the choice of r. We will later make use of this liberty in choosing r to obtain further information about the characteristic power series.

An interesting special case of the spaces $M(S)$ is the following: for each $\alpha \geq 0$ let $\{\lambda_1, \lambda_2, \ldots, \lambda_s\}$ be the set of eigenvalues of the U operator satisfying $\mathrm{ord}(\lambda_i) = \alpha$ (which is finite, as we have seen above), and let

$$g_\alpha = \prod_i (X - \lambda_i).$$

Then $M^{(\alpha)} = M(g_\alpha)$ is the space of all generalized eigenforms for U corresponding to eigenvalues of valuation α, which we call the "slope α eigenspace" of $\mathsf{M}(B, k, \mathrm{N}; 1) \otimes K$; we have obtained, for each α, a direct sum decomposition

$$\mathsf{M}(B, k, \mathrm{N}; 1) \otimes K = M^{(\alpha)} \oplus F^{(\alpha)}.$$

We denote the projection on $M^{(\alpha)}$ by e_α; as we have seen, $e_\alpha \in \mathbf{T}_{(k)} \otimes K$. Finally, we know that $M^{(\alpha)}$ is finite-dimensional and contained in the space $\mathsf{M}(B, k.\mathrm{N}; r) \otimes K$ of modular forms with growth condition r, for any r satisfying $\mathrm{ord}(r) < p/(p+1)$.

Similarly, one may define subspaces $M^{(\leq\alpha)}$ (respectively, $M^{(<\alpha)}$) corresponding to the generalized eigenforms with eigenvalues of valuation less than or equal to α (respectively, less than α), which are also finite-dimensional (again because $P(t)$ is entire), and for which we have continuous projections and direct sum decompositions as before.

It is clear, from the results of the preceding section, that one *cannot* extend these results to the full space $\mathsf{M}(B, k, \mathrm{N}; 1) \otimes K$. In fact, we have shown that the spectrum of U on this space contains the maximal ideal of B, so that the situation is dramatically different from what happens in the overconvergent case. It is interesting, in any case, to consider the largest subspace of $\mathsf{M}(B, k, \mathrm{N}; 1)$ on which U still acts "reasonably". This should be the union of all the overconvergent spaces, and that is essentially how we will define it. However, we have to be a little careful, because we have assumed B to be a discrete valuation ring and also that $r \in B$, so that we cannot make r tend to 1 unless we consider extensions of B.

Definition II.3.19 *Let B and* N *be as above. For any $f \in \mathsf{M}(B,k,\mathrm{N};1)$, we say f is overconvergent if there exist a positive integer m, a finite extension B_1 of B, and an element r in the maximal ideal of B_1 such that $p^m f \in \mathsf{M}(B,k,\mathrm{N};r)$. We denote the set of such f by $\mathsf{M}^\dagger(B,k,\mathrm{N};1)$, and give it the p-adic topology induced by that of $\mathsf{M}(B,k,\mathrm{N};1)$.*

We know that $\mathsf{M}^\dagger(B,k,\mathrm{N};1)$ is a dense subspace of $\mathsf{M}(B,k,\mathrm{N};1)$, since it contains all the finite sums

$$\sum b_a \mathrm{E}_{p-1}^{-a}.$$

The dagger notation is intended to recall the "weakly complete" spaces of Washnitzer and Monsky (see [MoWa68]); the analogy is that in both cases one considers the elements which can be written as power series with coefficients that tend to zero "better than linearly". To see this in our situation, one uses Corollary I.2.8, which shows that $f \in \mathsf{M}(B,k,\mathrm{N};1)$ will be overconvergent (with this definition) if and only if, when we write it as

$$f = \sum b_a \mathrm{E}_{p-1}^{-a},$$

there exists some rational number α such that $\mathrm{ord}(b_a) - a\alpha \to \infty$ as $a \to \infty$.

The interest of the space $\mathsf{M}^\dagger(B,k,\mathrm{N};1)$ is that the projections e_α are still defined on it, since any overconvergent form is contained in some space $\mathsf{M}(B,k,\mathrm{N};r) \otimes K$. They are, however, *not* continuous in the q-expansion topology when $\alpha \neq 0$ (because otherwise they would extend to the full space of p-adic modular forms). Thus, U is a "nuclear operator" on $\mathsf{M}^\dagger(B,k,\mathrm{N};1) \otimes K$, i.e., it has a spectral theory:

Corollary II.3.20 *Let $p \geq 7$, $\mathrm{N} \geq 3$, and assume $k \neq 1$ or $\mathrm{N} \leq 11$. Let $g(X)$ be a polynomial with nonzero independent term.*

Then we have a decomposition

$$\mathsf{M}^\dagger(B,k,\mathrm{N};1) \otimes K = M(g) \oplus F^\dagger(g)$$

such that $g(\mathrm{U})$ is nilpotent on $M(g)$ and invertible on $F^\dagger(g)$. The space $M(g)$ is finite-dimensional, coincides with the space $M(g)$ in Corollary II.3.18, and consists of the overconvergent modular forms of weight k on which $g(\mathrm{U})$ is nilpotent.

The characteristic power series $P(t)$ of the U *operator and the spaces $M(S)$ coincide with those in Corollary II.3.18. Finally, for each $\alpha \geq 0$, the set of eigenvalues λ satisfying $0 \leq \mathrm{ord}(\lambda) \leq \alpha$ is finite.*

Proof: This is all immediate from Corollary II.3.18 together with the fact that the characteristic power series is independent of r; see [Mons71] for a general statement on unions of nuclear spaces. □

II.3.5 Spectral theory: the ordinary case

The part of the spectral theory of U dealing with those eigenvalues which are units in the p-adic ring B is especially easy to understand. This seems to have been first noticed by Hida, who went on to construct a rich and powerful theory about this situation. In this section, we wish only to point out the most elementary fact about the ordinary case: that the ordinary projection e_0 does extend to the full space $\mathsf{M}(B,k,\mathrm{N};1)$. The full implications of this are drawn out in Hida's theory, the first steps of which we trace in an appendix to this chapter.

We use the notations introduced above; in particular,

$$e_0 : \mathsf{M}^\dagger(B,k,\mathrm{N};1)\otimes K \longrightarrow \mathsf{M}^\dagger(B,k,\mathrm{N};1)\otimes K$$

is the projection on the unit-root eigenspace, i.e., on the subspace generated by all generalized eigenforms whose eigenvalues are p-adic units, and $M^{(0)} = e_0\mathsf{M}^\dagger(B,k,\mathrm{N};1)$. We refer to e_0 as the "ordinary projection" and to $M^{(0)}$ as the "ordinary subspace".

To show that e_0 extends to $\mathsf{M}(B,k,\mathrm{N};1)$, we will in fact show a much stronger result, namely, that one can define an operator e_0 on all of $\mathbf{V}(B,\mathrm{N})$ which restricts to e_0 on $\mathsf{M}^\dagger(B,k,\mathrm{N};1)$. We do this, following Hida, by using the description of $\mathbf{V}(B,\mathrm{N})$ as the closure of the space of divided congruences $\mathsf{D}(B,\mathrm{N}p) = \varinjlim_k \mathsf{D}_k(B,\mathrm{N}p)$. To define e_0 as an operator on $\mathsf{D}(B,\mathrm{N}p)$, it is sufficient to define it on each $\mathsf{D}_k(B,\mathrm{N}p)$ in a coherent way; it will then extend to all of $\mathbf{V}(B,\mathrm{N})$ by continuity (because it is an endomorphism of $\mathsf{D}(B,\mathrm{N}p)$, and not merely of $\mathsf{D}(B,\mathrm{N}p)\otimes K$). However,

$$\mathsf{D}_k(B,\mathrm{N}p)\otimes K = \bigoplus_{i\le k} M(B,i,\mathrm{N}p)\otimes K \subset \bigoplus_{i\le k} \mathsf{M}^\dagger(B,i,\mathrm{N};1)\otimes K,$$

and we already know that e_0 is defined on this last space. Furthermore, after finite base-change, $\mathsf{D}_k(B,\mathrm{N}p)\otimes K$ has a basis $\{f_i\}$ which consists of eigenforms for the U operator (because $p \nmid \mathrm{N}$), and the action of f_i on this basis is simply given by $e_0 f_i = f_i$ if the eigenvalue of U on f_i is a unit, and $e_0 f_i = 0$ otherwise. Thus, e_0 maps $\mathsf{D}_k(B,\mathrm{N}p)\otimes K$ to itself.

It remains to show that $e_0\mathsf{D}_k(B,\mathrm{N}p) \subset \mathsf{D}_k(B,\mathrm{N}p)$; to see this, order the basis $\{f_i\}$ so that $e_0 f_i = f_i$ for $1 \le i \le r$ and $e_0 f_i = 0$ for $r+1 \le i \le m$. Suppose we have $f \in \mathsf{D}_k(B,\mathrm{N}p)$; then we can write

$$f = \sum_{1\le i\le m} \alpha_i f_i,$$

with $\alpha_i \in K$, and we know that

$$\sum_{1\le i\le m} \alpha_i f_i(q) \in B[[q]].$$

What we need to show is that this implies that in fact

$$\sum_{1\le i\le r} \alpha_i f_i(q) \in B[[q]].$$

To see this we note that U acts integrally on q-expansions, and that we have $\mathrm{U}^n f_i = \lambda_i^n f_i$, with λ_i a unit if $1 \le i \le r$ and in the maximal ideal otherwise. If we choose n large enough, we may assume that, first, $\alpha_i\lambda_i^n - \alpha_i \in B$ for $1 \le i \le r$, and second, that $\alpha_i\lambda_i^n \in B$ for the other i. Thus,

$$(e_0 f)(q) = \sum_{1\le i\le r} \alpha_i f_i(q) = (\mathrm{U}^n f)(q) - \sum_{1\le i\le r} (\lambda_i^n\alpha_i - \alpha_i) f_i(q) - \sum_{r+1\le i\le m} \alpha_i\lambda_i^n f_i(q) \in B[[q]].$$

Thus we have shown:

Lemma II.3.21 [**Hida**] *The operator e_0 induced on $\mathsf{D}_k(B, \mathrm{N}p) \otimes K$ from the operator e_0 on $\bigoplus \mathsf{M}^\dagger(B, i, \mathrm{N}; 1) \otimes K$ satisfies*

$$e_0\mathsf{D}_k(B, \mathrm{N}p) \subset \mathsf{D}_k(B, \mathrm{N}).$$

It is clear, then, that one can go to the inverse limit to get $e_0 : \mathsf{D}(B, \mathrm{N}) \longrightarrow \mathsf{D}(B, \mathrm{N})$, and then extend by continuity to $e_0 : \mathbf{V}(B, \mathrm{N}) \longrightarrow \mathbf{V}(B, \mathrm{N})$. In fact, the proof actually shows that e_0 restricted to $\mathsf{D}_k(B, \mathrm{N})$ can be expressed as a limit of powers of the U operator, and hence that in fact $e_0 \in \mathbf{T}(B, \mathrm{N})$, and in particular that e_0 commutes with the diamond operators, and hence preserve weights. Hence, finally, we may restrict to

$$e_0 : \mathsf{M}(B, k, \mathrm{N}; 1) \longrightarrow \mathsf{M}(B, k, \mathrm{N}; 1),$$

which by construction extends the e_0 defined on the overconvergent space.

Thus we get:

Proposition II.3.22 *Let $p \ge 7$, $\mathrm{N} \ge 3$, and assume $k \ne 1$ or $\mathrm{N} \le 11$. Then we have a decomposition*

$$\mathsf{M}(B, k, \mathrm{N}; 1) = e_0\mathsf{M}(B, k, \mathrm{N}; 1) \oplus (1 - e_0)\mathsf{M}(B, k, \mathrm{N}; 1).$$

The space $e_0\mathsf{M}(B, k, \mathrm{N}; 1)$ is finite-dimensional and we have

$$e_0\mathsf{M}(B, k, \mathrm{N}; 1) \subset \mathsf{M}(B, k, \mathrm{N}; r) \otimes K$$

for any r with $\mathrm{ord}(r) < p/(p+1)$. In particular, any ordinary eigenform is overconvergent.

Proof: All we need is to note that e_0 extends to $\mathsf{M}(B, k, \mathrm{N}; 1)$, and that, since $M^{(0)}$ is finite-dimensional, it is closed (with respect to any topology). □

In fact, Hida has shown much more; for example, for $k \ge 3$ any ordinary eigenform is in fact classical, rather than merely overconvergent. Furthermore, he has obtained a quite precise understanding of the Hecke algebra in the ordinary case. See the appendix for more information.

II.3.6 The characteristic power series

The characteristic power series of the U operator on the space of overconvergent p-adic modular forms of weight k is an intrinsically interesting object, and we would like to understand it better. In [Ka73], for example, Katz relates the reduction modulo p of this characteristic power series to the L-function of a certain algebraic variety.

In this section, we give somewhat explicit estimates for the coefficients of the characteristic power series. These give, for example, a lower bound for its Newton polygon, and imply several interesting results about congruences of modular forms.

Let us fix an integral weight $k \in \mathbf{Z}$, and denote by $P(t)$ the characteristic power series of U acting on $\mathsf{M}^\dagger(B,k,\mathrm{N};1) \otimes K$. In order to be able to use the full strength of the spectral theory, it is better to work with spaces of overconvergent forms; we may do this, since, as discussed above, $P(t)$ is also the characteristic power series for U acting on any space

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

where $r \in B$ is any element satisfying $\mathrm{ord}(r) < 1/(p+1)$, which we interpreted as the composition

$$\mathsf{M}(B,k,\mathrm{N};r^p) \otimes K \hookrightarrow \mathsf{M}(B,k,\mathrm{N};r) \otimes K \xrightarrow{u} \mathsf{M}(B,k,\mathrm{N};r^p) \otimes K,$$

where u is the map induced by the U operator on the full space. To obtain our estimates, we will in fact prefer to change our space (and our operator) slightly. We may continue the sequence of maps above:

$$\mathsf{M}(B,k,\mathrm{N};r^p) \otimes K \overset{i}{\hookrightarrow} \mathsf{M}(B,k,\mathrm{N};r) \otimes K \xrightarrow{u} \mathsf{M}(B,k,\mathrm{N};r^p) \otimes K \overset{i}{\hookrightarrow} \mathsf{M}(B,k,\mathrm{N};r) \otimes K$$

where we denote the inclusion by i; we have defined the U operator to be $u \circ i$. Note, however, that

$$\det(1 - t(u \circ i)) = \det(1 - t(i \circ u))$$

(see [Se62]), so that $P(t)$ is also the characteristic polynomial of the operator

$$i \circ u : \mathsf{M}(B,k,\mathrm{N};r) \otimes K \longrightarrow \mathsf{M}(B,k,\mathrm{N};r) \otimes K,$$

which we will also denote by U. (This is, of course, *also* the operator induced by the U operator on the full space!) This is the operator we will work with for our estimates.

We use a method suggested by Dwork in [Dw73]: using the "basis" for the space of p-adic modular forms constructed in Section I.2.2, we construct an explicit Banach basis for $\mathsf{M}(B,k,\mathrm{N};r) \otimes K$, and estimate the matrix coefficients of U with respect to this basis. As in Lemma II.3.9, the chosen basis defines a topology on $\mathsf{M}(B,k,\mathrm{N};r) \otimes K$ which is equivalent to the p-adic topology but with respect to which U is integral, i.e., $\|\mathrm{U}\| = 1$.

Let B be a finite extension of $\mathbf{Z}_p$, and suppose the level N and the weight k are fixed. Since the characteristic power series is independent of r (in the correct range), we may assume (provided $p \geq 7$) that

$$\frac{2}{3(p-1)} < \mathrm{ord}(r) < \frac{1}{p+1}.$$

For each i, choose a basis $\{b_{i,j},\ 1 \leq j \leq m_i\}$ of the space $A(B,k,i,\mathrm{N})$ (defined in Section I.2.2) which remains a basis after reduction modulo the maximal ideal[2] (for example, by choosing any lifting of a basis of $\mathsf{M}(\mathsf{k},k,\mathrm{N};1)$, where k is the residue field of B). We may assume that such bases have in fact been chosen consistently for all k and i, in the obvious sense (i.e., multiplication by E_{p-1} sends basis elements to basis elements). Then, for each m, the set

$$\{\mathrm{E}_{p-1}^{m-i} b_{i,j} \,|\, 1 \leq j \leq m_i,\ 0 \leq i \leq m\}$$

is a basis for the space $M(B, k+m(p-1), \mathrm{N})$ of classical modular forms of weight $k+m(p-1)$, so that we have:

$$\sum_{0 \leq i \leq m} m_i = \mathrm{rank}_B\, M(B,k+m(p-1),\mathrm{N}) = \dim M(K,k+m(p-1),\mathrm{N}),$$

which determines the dimensions m_i. (Note that the m_i are bounded independent of i.) Then, after Proposition I.2.6, we consider the Banach basis of $\mathsf{M}(B,k,\mathrm{N};r) \otimes K$ given by

$$e_{i,j} = r^i b_{i,j} \mathrm{E}_{p-1}^{-i}, \qquad i \geq 0,\ 1 \leq j \leq m_i.$$

This is clearly an orthonormal basis with respect to the standard p-adic topology on our space.

We want to modify the basis $e_{i,j}$ to obtain a basis defining an equivalent topology for which the U operator is integral. To see how this can be done, we first look at the matrix of U with respect to the basis $e_{i,j}$; write

$$\mathrm{U}(e_{i,j}) = \sum u_{i,j}^{t,s} e_{t,s},$$

with $u_{i,j}^{t,s} \in K$. To estimate $u_{i,j}^{t,s}$, recall that we have

$$\mathrm{U}(e_{i,j}) \in \frac{1}{p}\mathsf{M}(B,k,\mathrm{N},r^p),$$

so that we can write it in the form

$$\begin{aligned} \mathrm{U}(e_{i,j}) &= \frac{1}{p}\sum r^{pt} b'_t \mathrm{E}_{p-1}^{-t} \\ &= \sum \frac{r^{t(p-1)}}{p} r^t \left(\sum x_{t,s} b_{t,s}\right) \mathrm{E}_{p-1}^{-t} \\ &= \sum \frac{r^{t(p-1)}}{p} x_{t,s} e_{t,s}, \end{aligned}$$

[2] Such a basis is called an orthonormal basis, since, if $x = \sum x_j b_{i,j}$, we have

$$||x|| = \sup_j |x_j|,$$

where $||\cdot||$ is the p-adic norm on our space and $|\cdot|$ is the valuation norm on B (normalized, as usual, by $|p| = 1/p$).

so that we get

$$\begin{aligned}
\operatorname{ord}(u_{i,j}^{t,s}) &= \operatorname{ord}(r^{t(p-1)}) - \operatorname{ord}(p) + \operatorname{ord}(x_{t,s}) \\
&\geq t(p-1)\operatorname{ord}(r) - 1 \\
&> t(p-1)\frac{2}{3(p-1)} - 1 = \frac{2t}{3} - 1 = \frac{2t-3}{3}.
\end{aligned}$$

Thus, we already have $\operatorname{ord}(u_{i,j}^{t,s}) \geq 0$ whenever $t \geq 2$, and we need only modify our basis slightly to deal with the non-integrality of the other coefficients.

Define

$$\overline{e}_{i,j} = \begin{cases} \frac{1}{p} e_{i,j} & \text{if } i = 0 \\ \frac{1}{p^{1/3}} e_{i,j} & \text{if } i = 1 \\ e_{i,j} & \text{if } i \geq 2 \end{cases},$$

where we extend the ring B if necessary. Let $\overline{u}_{i,j}^{t,s}$ denote the matrix of U with respect to this basis. Then it is easy to see that

$$\operatorname{ord}(\overline{u}_{i,j}^{t,s}) \geq \begin{cases} 0 & \text{if } t = 0,\ 1 \leq s \leq m_0 \\ (p-1)\operatorname{ord}(r) - 2/3 > 0 & \text{if } t = 1,\ 1 \leq s \leq m_1 \\ t(p-1)\operatorname{ord}(r) - 1 > (2t-3)/3 & \text{if } t \geq 2,\ 1 \leq s \leq m_t \end{cases}$$

This, together with [Se62, Prop. 7], gives the following estimate for the coefficients of the characteristic power series: let

$$P(t) = \sum c_n t^n$$

be the characteristic power series of U on $\mathrm{M}(B, k, \mathrm{N}; r) \otimes K$; then we have

$$\operatorname{ord}(c_n) \geq \frac{m_2 + 3m_3 + 5m_4 + \ldots + (2(t-1)-3)m_{t-1}}{3},$$

where t is chosen so that

$$m_0 + m_1 + m_2 + \ldots + m_t \geq n > m_0 + m_1 + \ldots + m_{t-1}.$$

For a more precise statement, let

$$d_i = m_0 + m_1 + \cdots + m_i = \operatorname{rank}_{\mathbf{Z}_p} M(\mathbf{Z}_p, k + i(p-1), \mathrm{N}).$$

Then we can say:

Corollary II.3.23 *Assume $p \geq 7$, $\mathrm{N} \geq 3$, and either $k \neq 1$ or $\mathrm{N} \leq 11$. Let $P(t) = \sum c_n t^n$ be the characteristic power series of* U*; then we have:*

i. if $0 \leq n \leq d_1$, $\operatorname{ord}(c_n) \geq 0$,

ii. if $d_1 < n \leq d_2$,

$$\operatorname{ord}(c_n) \geq \frac{n - d_1}{3} \geq 0$$

iii. if $d_2 < n \leq d_3$,

$$\operatorname{ord}(c_n) \geq \frac{m_2 + 3(n - d_2)}{3} \geq \frac{m_2}{3}$$

iv. if $d_3 < n \leq d_4$

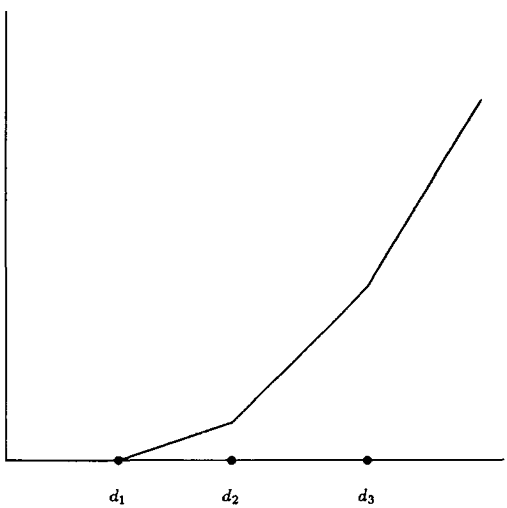
$$\operatorname{ord}(c_n) \geq \frac{m_2 + 3m_3 + 5(n - d_3)}{3} \geq \frac{m_2 + 3m_3}{3}$$

v. etc.

Proof: The result in [Se62] is that $\operatorname{ord}(c_n)$ is bounded by the sum of the bounds on the first n matrix coefficients, so that it is simply a matter of determining which are the first n estimates, in terms of the indices (t, s). Since the estimates depend only on the index t, this amounts to comparing n to the dimensions d_i. □

We can restate our result in terms of Newton polygons, which makes it much easier to grasp:

Corollary II.3.24 *The Newton polygon of* $P(t)$ *lies above the polygon*



which has

i. slope 0 *for* $0 < x < d_1$

ii. slope $1/3$ *for* $d_1 < x < d_2$

iii. slope $1 = 3/3$ *for* $d_2 < x < d_3$

iv. slope $5/3$ *for* $d_3 < x < d_4$

v. etc.

As an application of this analysis, consider the following problem: it follows from the work of Jochnowitz (see [Jo82b]) that any eigenform for the U operator which is "ordinary", i.e., has unit eigenvalue, must be congruent (modulo the maximal ideal of B) to a classical modular form of weight at most $p+1$; is it possible to give a similar bound when the eigenvalue is of valuation α? (Note that the point here is the bound on the weight, since any p-adic modular form will be congruent to some classical modular form, almost by definition.) It turns out that our analysis gives just such a bound. Our result will be weaker than that of Jochnowitz, however, on two counts: first, her result applies to generalized eigenforms, i.e., to any f in the slope zero eigenspace denoted above by $M^{(0)}$, and second, because any modular form which is congruent to an ordinary form must itself be ordinary, and this is unfortunately not the case for forms of higher slope.

Thus, let f be an eigenform for the U operator, of slope α, so that we have $\mathrm{U}(f) = \lambda f$, and $\mathrm{ord}(\lambda) = \alpha$. As we have shown above, f must be overconvergent, so we may assume $f \in \mathsf{M}(B, k, \mathrm{N}; r)$, with r as above. Write f in terms of the first basis constructed above, so that

$$f = \sum f_{i,j} e_{i,j}.$$

Then, with notations as above, we have:

$$\begin{aligned} \lambda \sum f_{i,j} e_{i,j} = \lambda f = \mathrm{U} f &= \sum f_{i,j} \sum \frac{r^{t(p-1)}}{p} x_{t,s} e_{t,s} \\ &= \sum \frac{r^{i(p-1)}}{p} f'_{i,j} e_{i,j}. \end{aligned}$$

Equating the valuations of the coefficients, we get

$$\begin{aligned} \mathrm{ord}(\lambda) + \mathrm{ord}(f_{i,j}) &= i(p-1)\,\mathrm{ord}(r) - 1 + \mathrm{ord}(f'_{i,j}) \\ &\geq i(p-1)\,\mathrm{ord}(r) - 1 \\ &> \frac{2i-3}{3}. \end{aligned}$$

Hence, if

$$i \geq \frac{3}{2}(\mathrm{ord}(\lambda) + 1),$$

we get

$$\mathrm{ord}(\lambda) + \mathrm{ord}(f_{i,j}) > \mathrm{ord}(\lambda) + 1 - 1 = \mathrm{ord}(\lambda),$$

so that

$$\mathrm{ord}(f_{i,j}) > 0.$$

Let

$$n(\alpha) = \lfloor \frac{3}{2}(\alpha + 1) \rfloor,$$

where $\lfloor\cdot\rfloor$ is the greatest integer function. Then, recalling that $\mathrm{ord}(\lambda)=\alpha$, we conclude that

$$f = g\mathrm{E}_{p-1}^{-n(\alpha)} + h,$$

where g is a classical modular form of weight $k+n(\alpha)(p-1)$ and $h \in \mathfrak{m}\mathsf{M}(B,k,\mathrm{N};r)$ is congruent to zero modulo the maximal ideal $\mathfrak{m}$ of B. Hence, we have shown:

Proposition II.3.25 *Let $f \in \mathsf{M}(B,k,\mathrm{N};r)$ satisfy $\mathrm{U}f=\lambda f$, with $\mathrm{ord}(\lambda)=\alpha$, and let*

$$n(\alpha) = \lfloor \frac{3}{2}(\alpha+1)\rfloor.$$

Then there exists a classical modular form g of level N and weight $k+n(\alpha)(p-1)$ such that

$$f \equiv g\mathrm{E}_{p-1}^{-n(\alpha)} \pmod{\mathfrak{m}}$$

in $\mathsf{M}(B,k,\mathrm{N};r)$.

Note that congruence modulo $\mathfrak{m}\mathsf{M}(B,k,\mathrm{N};r)$ is a much stronger fact than congruence in q-expansion.

This result can be improved in several ways; for example, one should remark that the proof will go through if we simply assume that

$$\mathrm{U}f \equiv \lambda f \pmod{p^{\lfloor\alpha\rfloor+1}}.$$

This allows us to show:

Corollary II.3.26 *There exists a constant $C(\alpha)$, depending on α but not on k or N, such that for $k \geq C(\alpha)$ any eigenform $f \in \mathsf{M}(B,k,\mathrm{N};r)$ satisfying $\mathrm{U}f=\lambda f$ with $\mathrm{ord}(\lambda)=\alpha$ is congruent modulo $\mathfrak{m}$ to a classical eigenform of the* same *weight k.*

For the case $\alpha=0$, the bound obtained in this way is $C(0)=p-1$, which is of course far weaker than that obtained Jochnowitz, which is $C(0)=3$. As Hida shows in [Hi86b], Jochnowitz's bound implies that any ordinary eigenform of sufficiently high weight *is* classical. The same would follow in the general case if we could show that the classical eigenform whose existence is asserted is also of slope α. Such a result, if true, would be closely connected with the questions of interpolation discussed in the next section.

II.3.7 Varying the weight

From the point of view of the general theory of p-adic modular functions, we would like to "put together" the projections e_α for varying weights in order to get a projection defined on some subspace of $\mathbf{V}[1/p]=\mathbf{V}(\mathbf{Z}_p,\mathrm{N})\otimes\mathbf{Q}_p$ (the subspace of p-adic modular functions that are "overconvergent", in some sense). For this, one needs to analyze the dependence of $P(t)$ and of the projections e_α on the weight k and the parameter r. (Of course, $P(t)$ does not depend on r, but the norm of e_α might very well do so.) What one expects is that both these constructions should vary analytically with the weight, at least within a restricted domain. In the case of e_α, we would at least like to obtain a uniform bound for the norm when r is fixed and the weight varies. No results of this kind are known, except in the case $\alpha = 0$, which has been considered above. In the general case, such results turn out to be quite elusive, and we will limit ourselves, in this section, to formulating a few conjectures and pointing out their importance.

Let $P_k(t)$ denote the characteristic power series of the U operator on the space $\mathsf{M}^\dagger(B,k,\mathrm{N};1)\otimes K$, obtained as above. Then we have

$$P_k(t) = \prod_i (1-\lambda_i^{(k)}t),$$

where the $\lambda_i^{(k)}$ are the eigenvalues of U on $\mathsf{M}^\dagger(B,k,\mathrm{N};1)\otimes K$, so that we know that the $\lambda_i^{(k)}$ are all integral and form a sequence tending to zero. Hence, to determine $P_k(t) \pmod{p^n}$, it is enough to know the eigenvalues of valuation less than n, together with their multiplicities. However, since reducing modulo p^n may introduce extraneous eigenvalues, this is a quite subtle problem, and we will offer only a few remarks about it and hints as to how one might proceed.

The first natural conjecture is that the characteristic power series varies continuously as one varies the weight. This amounts to the following conjecture:

Conjecture II.1 *Suppose $k_2 = k_1 + ip^{n-1}(p-1)$. Then we have $P_{k_1}(t) \equiv P_{k_2}(t) \pmod{p^n}$.*

To see that this is indeed plausible, note that if $k_2 = k_1 + ip^{n-1}(p-1)$ for some $i \in \mathbf{Z}$, multiplication by $\mathrm{E}_{p-1}^{ip^{n-1}}$ gives an isomorphism

$$\mathsf{M}^\dagger(B,k_1,\mathrm{N};1) \xrightarrow{\mathrm{E}_{p-1}^{ip^{n-1}}} \mathsf{M}^\dagger(B,k_2,\mathrm{N};1)$$

which is U-equivariant mod p^n (because the q-expansion of $\mathrm{E}_{p-1}^{p^{n-1}}$ is congruent to 1 mod p^n). This suggests that the characteristic power series will then be necessarily congruent modulo p^n, but does not furnish a proof.

The continuity of the map $k \mapsto P_k(t)$ would already be a significant result. For example, let

$$M_k^{(\alpha)} \subset \mathsf{M}^\dagger(B,k,\mathrm{N};1)\otimes K$$

denote the slope α subspace of the space of overconvergent p-adic modular forms of weight k. If we assume that Conjecture II.1 is true, it follows, by considering Newton polygons and their interpretation in terms of the number (counting multiplicity) of roots of a certain valuation, and hence of the dimension of the slope α eigenspaces, that the map

$$k \mapsto d(k,\alpha) = \dim M_k^{(\alpha)}$$

is *locally constant*. By compactness, it follows that it is uniformly locally constant, so that we would get:

Conjecture II.2 *For every $\alpha \geq 0$, there exists a positive integer $m(\alpha) \in \mathbf{Z}$ such that, whenever $k_1 \equiv k_2 \pmod{p^{m(\alpha)}(p-1)}$, we have $d(k_1, \alpha) = d(k_2, \alpha)$.*

Once again, this is known if $\alpha = 0$, in which case we can take $m(\alpha) = 0$ (see the appendix to this chapter). In general, we suspect that $m(\alpha)$ and α should have the same order of magnitude, but we have no real evidence for a conjecture.

Given the conjecture that the characteristic power series varies continuously with the weight, one may go further and ask if it is not in fact analytic (especially given Hida's theory of the ordinary part). We suspect that this is not the case. We will, however, formulate (very tentatively) a conjecture as to what is in fact the case. With notations as above, let

$$P_k^{(\alpha)}(t) = \prod_{\mathrm{ord}(\lambda_i^{(k)}) \leq \alpha} (1 - \lambda_i^{(k)} t)$$

be the "slope at most α" part of the characteristic power series. Note that each $P_k^{(\alpha)}(t)$ is a polynomial and that

$$\lim_{\alpha \to \infty} P_k^{(\alpha)}(t) = P(t).$$

Then it seems reasonable to make the following guess:

Conjecture II.3 *The map $k \mapsto P_k^{(\alpha)}(t)$ is locally analytic.*

In addition, one would expect "locally" to depend on α in such a way that the limit $P(t)$ is *not* itself analytic. For $\alpha = 0$, one would expect that, for each j, $0 \leq j < p-1$, there exists $P_j^{(0)} \in \Lambda[t]$ specializing to $P_k^{(0)}(t)$ for each $k \equiv j \pmod{p-1}$. That this is in fact the case follows easily from Hida's theory of the ordinary projection.

The other important question in relation to variation with the weight has to do with the projections e_α to the slope α part, which we defined above. The question is whether the e_α "compile well" as one varies the weight. For this to make sense, one must fix the parameter r.

Conjecture II.4 *The projections*

$$e_\alpha : \mathsf{M}(B, k, \mathrm{N}; r) \otimes K \longrightarrow \mathsf{M}(B, k, \mathrm{N}; r) \otimes K$$

are bounded independent of k, i.e., there exists $C(\alpha, r) \in \mathbf{R}$ such that we have $\|e_\alpha\| \leq C(\alpha, r)$ for any weight k, where we take the operator norm with respect to the p-adic topology on $\mathsf{M}(B, k, \mathrm{N}; r) \otimes K$.

It is easy to see, for example, that $\|e_0\| \leq 1$ independent of k, even in the q-expansion topology; this is what makes Hida's theory work. It is also easy to see that the projections are *not* integral when $\alpha > 0$. Of course, we have shown that they are bounded for each k, but our proof does not seem to provide either explicit bounds or estimates on how the bound varies with k. This remains an interesting open question.

II.4 Appendix: Hida's theory of the ordinary part

In this appendix, we give a short introduction to Hida's theory of the ordinary part. The crucial fact here, as we have already pointed out, is that we have $\|e_0\| \leq 1$ *in the q-expansion topology*, so that in fact we have a projection on $\mathbf{V}(B, \mathrm{N})$:

Proposition II.4.1 *There exists an idempotent $e_0 \in \mathbf{T}(\mathbf{Z}_p, \mathrm{N})$ giving a projection $e_0 : \mathbf{V}(B, \mathrm{N}) \longrightarrow \mathbf{V}(B, \mathrm{N})$ which induces the projection on the part of slope zero in each of the subspaces $\mathsf{M}^{\dagger}(B, k, \mathrm{N}; 1)$.*

Cutting everything down by e_0 defines the ordinary part:

Definition II.4.2 *The ordinary part of $\mathbf{V}$ is the subring*

$$\mathbf{V}^{ord} = e_0 \mathbf{V};$$

the ordinary Hecke algebra $\mathbf{T}^{ord}$ is the corresponding Hecke algebra, so that $\mathbf{T}^{ord} = e_0 \mathbf{T}$. Analogously, we define $\mathbf{V}^{ord}(B, \mathrm{N})$ and $\mathbf{T}^{ord}(B, \mathrm{N})$.

The diamond operators make $\mathbf{T}^{ord}$ a $\mathbf{\Lambda}$-algebra, where

$$\mathbf{\Lambda} = \mathbf{Z}_p[[\mathbf{Z}_p^{\times}]]$$

is the completed group ring. Splitting $\mathbf{\Lambda}$ into a sum of local rings gives

$$\mathbf{\Lambda} = \bigoplus_i \Lambda_{(i)},$$

where $(\mathbf{Z}/p\mathbf{Z})^{\times} \subset \mathbf{Z}_p^{\times}$ acts via the i-th power of the Teichmüller character on $\Lambda_{(i)}$. Let $\mathsf{T}_i = \mathbf{T}^{ord} \otimes_{\mathbf{\Lambda}} \Lambda_{(i)}$ be the corresponding decomposition of the ordinary Hecke algebra. The first crucial result is then the following:

Theorem II.4.3 [Hida] *With the above definitions, we have:*

i. *$\mathsf{T}_i = \mathsf{T}_i(\mathbf{Z}_p, \mathrm{N})$ is a finite flat Λ-algebra of rank $r(i)$ given by*

$$r(i) = \mathrm{rank}_{\mathbf{Z}_p} e_0 M(\mathbf{Z}_p, k, \mathrm{N}) = \dim_{\mathbf{F}_p} e_0 M(\mathbf{F}_p, k, \mathrm{N}),$$

for any *k satisfying $k \geq 3$ and $k \equiv i \pmod{p-1}$;*

ii. *for any $k \geq 3$, the ordinary subspace*

$$M_k^{(0)} = e_0 \mathsf{M}(\mathbf{Z}_p, k, \mathrm{N}; 1) \subset \mathsf{M}(\mathbf{Z}_p, k, \mathrm{N}; 1)$$

consists of ordinary projections of classical modular forms of weight k and level N, i.e.,

$$M_k^{(0)} \subset e_0 M(\mathbf{Z}_p, k, \mathrm{N}) \subset M(\mathbf{Z}_p, k, \Gamma_1(\mathrm{N}) \cap \Gamma_0(p)).$$

Proof: See [Hi86b], but note that, given the duality theory developed in the next chapter (or Hida's analogous results), this follows easily from the work of Jochnowitz referred to above (see [Jo82b]). □

This theorem is the starting point of Hida's work, which connects to it in several different ways. In [Hi86b], Hida has investigated the structure of the Hecke algebra T_i, relating it to the existence of congruences between systems of eigenforms and to Iwasawa theory. In [Hi86a], he uses this to construct families of Galois representations, which we discuss ahead. We refer to Hida's papers (for example, [Hi86b] and [Hi86a]) for more details.