# Lecture Notes in Mathematics

Edited by A. Dold and B. Eckmann

1304

Fernando Q. Gouvêa

# Arithmetic of *p*-adic Modular Forms

Springer-Verlag
Berlin Heidelberg New York London Paris Tokyo

**Author**

Fernando Quadros Gouvêa
Instituto de Matemática e Estatística
Universidade de São Paulo
Caixa Postal 20570 (Ag. Iguatemi)
01498 São Paulo, SP, Brazil

Mathematics Subject Classification (1980, revised 1985): 11 F

ISBN 3-540-18946-7 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-18946-7 Springer-Verlag New York Berlin Heidelberg



Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.
2146/3140-543210

# Introduction

The theory of $p$-adic modular forms was initiated by Serre, Katz, and Dwork, who, in the early 1970's, attempted to define objects which would be recognizably modular forms but which would be truly $p$-adic, reflecting the $p$-adic topology in an essential way. Specifically, one wanted a theory where two modular forms with highly congruent $q$-expansion coefficients would be $p$-adically close, and where limits of modular forms (with respect to this topology) would exist. The initial motivation for this construction was the problem of $p$-adic interpolation of special values of L-functions.

The first difficulty, of course, was to find the correct definitions. Serre's approach (see [Se73]) was the most elementary: modular forms were identified with their $q$-expansions and $p$-adic modular forms were considered as limits of such $q$-expansions. Serre showed that such limits have "weights" (in a suitable sense), and developed the theory sufficiently to be able to obtain $p$-adic L-functions by constructing suitable analytic families of $p$-adic modular forms. His theory of analytic families of modular forms will be briefly discussed in our third chapter. He was also the first to notice that modular forms of level $Np$ (and appropriate nebentypus) are of level N when considered as $p$-adic modular forms, which is a crucial aspect of the $p$-adic theory.

Dwork's approach was more analytic. In [Dw73], for example, he restricts himself to $p$-adic analytic functions on a modular curve (i.e., $p$-adic modular forms of weight zero); in the same article, he notes the importance of growth conditions and of the fact that the U operator is completely continuous (in the sense of Serre) in the case which he is considering.

Katz's work brought these approaches together, showing how to define $p$-adic modular forms in modular terms, generalizing the results of Dwork on the U operator and of Serre on congruences ([Ka73], [Ka75a], and [Ka76]). As defined by Katz, $p$-adic modular forms are clearly *modular* objects, i.e., they are clearly related to classifying elliptic curves with some additional structure. This allows a much more conceptual approach to the theory, but at the same time makes it less accessible. Using his results, Katz was able to obtain important results on higher congruences between $q$-expansions of modular forms and several interpolation theorems for L-functions and Eisenstein series. It is this approach that we follow on foundational matters, and it will be fully described in our first chapter.

---

During the preparation of this book, the author received the financial support of the Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil, of the CAPES-COFECUB project, and of the Institut des Hautes Études Scientifiques in Bures-sur-Yvette, France.

More recently, attention has once again been called to the subject of $p$-adic modular forms by the work of Hida, who has used the theory, especially through his construction of the "ordinary part" of the space of $p$-adic modular functions, in several different contexts (see [Hi86b], [Hi86a], etc.). Most notably, from our point of view, he has shown how to use ordinary $p$-adic modular forms to construct analytic families of Galois representations. This has been considered in more detail by Mazur and Wiles in [MW86], and has motivated Mazur's construction of a general theory of deformations of Galois representations in [Ma], in relation to which several interesting results have been obtained by Mazur and Boston.

The aim of this book is twofold. In the first place, we have tried to put together a unified and coherent exposition of the foundations of the theory, pointing out, in particular, the connections between the various approaches of Serre, Katz, and Hida. In the course of so doing, we have filled in various gaps and obtained several new results. In particular, we have obtained quite a bit of information about the action of the U operator, especially as relates to the spaces of overconvergent forms. In the second place, we have considered the problem of constructing deformations of residual representations, and shown that many of these are attached (as in the classical case) to $p$-adic modular forms which are eigenforms under the action of the Hecke operators, provided the original residual representation is absolutely irreducible and attached to a classical modular form over a finite field.

The chapters break down as follows: the first chapter is largely foundational in nature, and most of its contents are to be found, implicitly or explicitly, in the works of Serre and Katz. It is intended to give the reader ready access to the theory. We have tried to give precise references for all the quoted theorems, and have frequently given informal descriptions of their proofs.

The second chapter introduces the Hecke algebra, and, in particular, the U operator, and considers the problem of obtaining eigenforms for the Hecke algebra by way of studying the spectral theory of the U operator. These results are extensions of those obtained by Katz in [Ka73] (he considers only the case of overconvergent modular forms of weight zero). The main payoff here is the fact that there are "very few" overconvergent eigenforms for the U operator outside the kernel of U, in the precise sense that the space of overconvergent eigenforms for U with fixed weight $k \in \mathbf{Z}$ and fixed valuation for the eigenvalue of U is finite-dimensional. In fact, if one requires that the eigenvalue of U be a unit (the "ordinary" case considered by Hida), one knows that any such eigenform of weight $k \geq 3$ will necessarily be classical. In the non-ordinary case (i.e., when the eigenvalue is assumed to belong to the maximal ideal of $B$), one does not know any examples that are not classical modular forms. By contrast, if we do not require overconvergence, one can produce a large number of (non-ordinary) eigenforms by a simple construction.

The third chapter deals with constructing Galois representations; for this, we first construct a good duality theory for the Hecke algebra (related to the one obtained by Hida), and then use it in a crucial way to construct a "universal modular deformation" of a given (modular) absolutely irreducible residual representation, which defines a subspace of Mazur's "universal deformation space". We then study this "universal modular deformation space" with a view to the natural question of determining its dimension, and, in particular, of deciding how close it comes to filling the entire deformation space. In this direction, we show that the Krull dimension of the modular deformation ring is at least three, which implies that the dimension of the modular deformation space, as a formal scheme, is at least two. One suspects that all of the deformations are in fact modular, so that the two deformation spaces are the same; we have not been able to show this, even in the well-understood "neat $S_3$" case considered by Mazur and Boston, but we have several partial results that tend to support the conjecture.

Readers who are already familiar with the theory of $p$-adic modular forms as formulated by Katz may wish to skip all of the first chapter and the first sections of the second, beginning to read at the point where we begin to study the U operator and its spectral theory. Readers who are only interested in Galois representations can begin directly with Chapter 3, and refer back as necessary.

Most of the work contained herein was done while at Harvard University and at the Institut des Hautes Études Scientifiques; I would like to thank both institutions for their hospitality, and in particular for having been able to use their computer facilities so as to give this text the appearance it has.

My debt to the mathematical work of Barry Mazur, Nicholas M. Katz, J.-P. Serre, and Haruzo Hida, among others, will, I hope, become abundantly clear in the text. On a more personal level, many people have helped me on the way to the completion of this book, which is a expanded and corrected version of my Harvard Ph.D. thesis. Barry Mazur, who was my thesis advisor, suggested the topic, expressed interest at every stage, and supplied crucial help at many points. David Roberts, Dinesh Thakur, Jeremy Teitelbaum, Alan Fekete, and others of my fellow graduate students at Harvard were always ready to discuss this and other topics, and contributed greatly to the intellectual excitement of doing mathematics. J. F. Mestre of the École Normale Superieure helped me with the complexities of incompatible computer equipment at a critical point in the preparation of the final version. My sons Heitor and Marcos reminded me every night that there was more to life than mathematics, and cheered "Daddy's big project" on from day to day. My parents helped in many ways, both financially and personally, and my wife contributed in more ways than one can count. I thank you all; may God bless you all greatly.

Bures-sur-Yvette, January of 1988

Fernando Quadros Gouvêa

...τι ποιεῖτε, πάντα εις δόξαν Θεοῦ ποιεῖτε.

# Contents

# Chapter I

# p-adic Modular Forms

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

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

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

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

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

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

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

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

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

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

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

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

# I.1 Level Structures and Trivializations

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

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

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

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

*of finite flat group schemes over B.*

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

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

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

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

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

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

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

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

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

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

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

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

is the canonical inclusion.

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

## I.2 $p$-adic Modular Forms with Growth Conditions

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

## I.2.1 Definitions

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

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

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

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

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

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

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

*satisfying the following conditions:*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## I.2.2 Basic Properties

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

such that

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

We also set

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

(which is just 0 when $k$ is negative) and

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

Thus we have isomorphisms

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

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

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

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

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

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

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

*induces, for any $r$, an isomorphism*

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

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

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

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

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

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

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

*Proof*: see [Ka73, Section 2.6]. □

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

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

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

*defined by transposition from the map of functors*

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

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

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

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

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

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

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

*Proof*: immediate from the previous corollary. □

Note in particular that

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

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

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

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

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

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

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

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

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

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

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

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

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

*Proof*: see [Ka73, Prop. 2.7]. □

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

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

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

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

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

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

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

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

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

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

**Corollary I.2.13** *Multiplication by* $\mathrm{E}_{p-1}$ *gives an isomorphism*

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

*The corresponding map on the "bases" is given by*

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

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

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

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

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

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

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

# I.3 Generalized p-adic Modular Functions

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

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

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

## I.3.1 Definition

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

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

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

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

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

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

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

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

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

so that

$$\mathbf{W} = \varprojlim_{n} \varinjlim_{m} \mathbf{W}_{n,m}. \tag{I.7}$$

Note that this definition is independent of the exponent $\nu$, so that we have

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

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

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

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

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

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

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

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

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

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

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

$$\mathbf{V}=\varprojlim_n\varinjlim_m\mathbf{V}_{n,m}, \tag{I.8}$$

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

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

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

$$\mathbf{V}_{par}=\varprojlim_n\varinjlim_m\mathbf{V}_{n,m}^{cusp}.$$

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

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

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

## I.3.2 The $q$-expansion map

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

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

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

*is injective, and the cokernel*

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

*is flat over $B$.*

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

Another useful result is

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

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

*Proof*: see [Ka76, Chapter 5]. □

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

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

## I.3.3 Diamond operators

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

Let

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

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

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

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

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

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

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

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

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

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

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

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

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

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

### I.3.4 Weight and nebentypus

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

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

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

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

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

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

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

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

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

### I.3.5 Modular forms and modular functions

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

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

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

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

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

and hence a homomorphism of rings

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

given by

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

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

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

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

defines maps

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

### I.3.6 Divided congruences

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

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

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

and then define the *ring of divided congruences* by

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

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

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

we have that

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

For example,

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

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

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

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

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

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

**Lemma I.3.8** *The map $\alpha_1$ sends* $\sum M(B, i, \mathrm{N})$ onto $\mathbf{V}_{1,1}$, *with kernel equal to the ideal generated by* $\mathrm{E}_{p-1} - 1$, *and hence gives an isomorphism*

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

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

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

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

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

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

and

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

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

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

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

[1]By analogy, one might consider the spaces

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

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

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

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

---

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

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

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

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

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

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

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

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

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

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

Passing to the limit, and since $\mathsf{S}$ is an ideal of $\mathsf{D}$, we get a surjection

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

which proves the proposition. □

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

## I.3.7 Appendix: $p$-adic modular forms of weight $\chi$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Since for every $n$ there are maps

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

taking the inverse limit gives a map

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

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

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

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

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

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

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

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

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

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

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

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

# Chapter II

# The Hecke and U Operators

In this chapter we define $p$-adic versions of the classical Hecke operators. For the operators $T_\ell$ with $\ell \neq p$, this is quite easy, and may be done either by imitating the classical definition in terms of subgroups of order $\ell$ 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

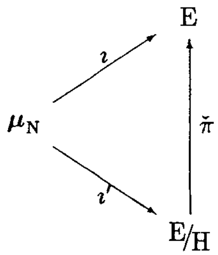
The goal of this section is to define Hecke operators $T_\ell$ on the ring $\mathbf{V}$ of generalized $p$-adic modular functions. One can do this either by giving an intrinsic modular definition or by noticing that one can define Hecke operators on $\mathbb{D}$ from the classical Hecke operators by 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

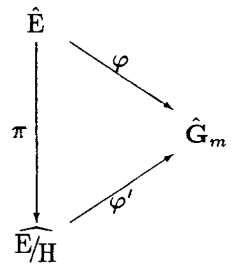
Let E be an elliptic curve over a $p$-adic ring A, $\varphi$ be a trivialization, and $\imath$ a $\Gamma_1(\mathrm{N}p^\nu)^{arith}$-structure on E. Let $\ell$ be a rational prime, $\ell \neq p$, not dividing N. For any subgroup H of order $\ell$ in E, we can consider the quotient curve ${}^{\mathrm{E}}\!/_{\mathrm{H}}$;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

$$\mathrm{E} \xrightarrow{\ \pi\ } {}^{\mathrm{E}}\!/_{\mathrm{H}},$$

and let $\check{\pi}$ denote the dual isogeny. Since $\ell$ does not divide $\mathrm{N}p^\nu$, both $\pi$ 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 ${}^{\mathrm{E}}\!/_{\mathrm{H}}$ by $\imath' = \check{\pi}^{-1} \circ \imath$.




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




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 $\mathrm{T}_\ell$ 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 $\epsilon$, 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 $\mathrm{T}_\ell$ as defined above.

To define Hecke operators $\mathrm{T}_\ell$ with $\ell$ dividing N (but different from $p$), we follow an analogous procedure, but sum only over those subgroups of order $\ell$ 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 $\mathrm{T}_\ell$ on $\mathbf{V}$ for $\ell|\mathrm{N}$, whose effect on $q$-expansions corresponds to that of the "$\mathrm{U}_\ell$" 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 $\mathrm{T}_p$). 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 $\mathrm{T}_\ell$ (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 $\mathrm{T}_\ell$ with $\ell$ 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 $\mathbf{T}^*$, which will be the completion of the algebra of endomorphisms of $\mathbf{V}$ generated by the diamond operators and the $T_\ell$ with $\ell$ not dividing $Np$. 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 $\mathbf{T}_0^*$. 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 $\mathsf{D}' = \mathsf{D}'(B, Np^\nu)$ of divided congruences of modular forms of level $Np^\nu$ (where, for this construction, we must assume $\nu \geq 1$, in order to have an action of U). We first note that, the operators $T_\ell$ for $\ell \neq p$ and U act on classical modular forms of weight $j \geq 1$ and level $Np^\nu$ and preserve congruences of modular forms over $\mathbf{Z}_p$, so that they act on the spaces $\mathsf{D}'_k$. We define the Hecke algebra of $\mathsf{D}'_k$, denoted by $\mathcal{H}'_k$, as the $\mathbf{Z}_p$-algebra of endomorphisms of $\mathsf{D}'_k$ generated by the endomorphisms induced by the $T_\ell$, by U, and by the diamond operators. As before, we also define $\mathcal{H}'^*_k$, by excluding the operator U and the $T_\ell$ with $\ell | N$. Then it is clear that we have restriction maps

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

and

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

whenever $j < k$, and the inverse limits

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

and

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

are $p$-adically complete algebras of continuous endomorphisms of $\mathsf{D}'$, which are uniformly continuous in the $q$-expansion topology (i.e., the topology induced from $\mathbf{V}$). Since $\mathsf{D}'$ is dense in $\mathbf{V}$, the actions of $\mathbf{T}$ and $\mathbf{T}^*$ extend to $\mathbf{V}$, and the action thus obtained coincides with the one we defined before (because it does so on the dense subspace $\mathsf{D}'$—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 $\mathsf{D}'$ rather than $\mathsf{D}$, 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 $\mathsf{D}$ or $\mathsf{D}'$.

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 $\mathrm{T}_p$). 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

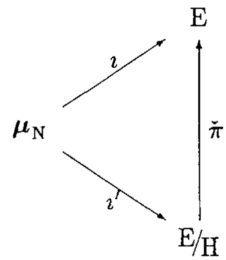
$$\sum a_n q^n \mapsto \sum a_n q^{np}.$$

This maps classical modular forms of level N to modular forms of level N$p$. Since modular forms of level $\mathrm{N}p^{\nu}$ 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.

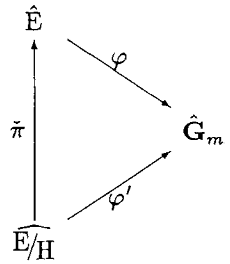
We define the Frobenius operator on the ring $\mathbf{V}$ by using the fact that the trivialization determines a canonical subgroup of order $p$ in E.

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

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




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




We have then defined a map of functors

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

which defines a map, the *Frobenius endomorphism* of **W**,

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

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

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

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

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

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 $\mathrm{ord}$ is normalized by $\mathrm{ord}(p) = 1$. To every test object $(\mathrm{E}/A, \imath, Y)$ of level $\mathrm{N}$ 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 $\mathrm{E}$, satisfying:*

- *$\mathrm{H}$ depends only on the isomorphism class of $(\mathrm{E}/A, Y)$,*
- *the formation of $\mathrm{H}$ commutes with arbitrary base change of $p$-adically complete $B$-algebras,*
- *if $p/r = 0$ in $A$, $\mathrm{H}$ is the kernel of the Frobenius map $\mathrm{E} \longrightarrow \mathrm{E}^{(p)}$,*
- *if $\mathrm{E}/A$ is the Tate curve $\mathrm{Tate(q)}$ over $(A/p^n A)((q))$, the fundamental subgroup $\mathrm{H}$ 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 $p^\nu$ for every $\nu$. 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* $p^\nu$ *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 **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 $\mathrm{F}^n$ 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* $p^\nu$*, called the fundamental subgroup of order* $p^\nu$ *of* E, *satisfying:*

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

*Proof*: We use induction on $\nu$. 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

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

we get, by induction, a fundamental subgroup $H' \subset E'$ of order $\nu - 1$. Let $E'' = E'/H'$, and let $f_\nu$ 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 $p^\nu$ 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 $p^\nu$ (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 $\operatorname{ord}(E_{p-1})$ 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 $\operatorname{ord}(E_{p-1})$, 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* $\operatorname{ord}(r) < 1/p^{\nu-2}(p+1)$.

*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 $\nu$) 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 $\nu$ 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 $\nu$), 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 $\mathrm{ord}(r) < p/(p+1)$. 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 $\lambda$ for *every* element $\lambda$ 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 $\alpha$ 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 $\lambda$, 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(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 $\chi$ (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* $p \nmid \mathrm{N}$. *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 $Tr_{\mathrm{Frob}}$ 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* $p \nmid \mathrm{N}$. *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 $\Omega$ be the completion of the algebraic closure of $K$, and let $B_\infty$ be its ring of integers. Let $\mathrm{E}/B_\infty$ be a supersingular elliptic curve over $B_\infty$, and let T be a parameter for the formal group of E, normalized by $[\zeta](T) = \zeta T$ for every $(p-1)^{\text{st}}$ root of unity $\zeta$ 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 $i \not\equiv 1(\mathrm{mod} p)$, and $\operatorname{ord}(c_p) = 0$ (because the formal group is of height 2). Note that since $a \equiv \mathrm{E}_{p-1}(\mathrm{E},\omega)$ (mod p) for any nonvanishing differential $\omega$ 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 $\mathrm{H}_1$, $\mathrm{H}_2$, ..., $\mathrm{H}_p$ 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 $B_\infty$, 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 $\mathrm{E}_{p-1}$.

**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 $\omega_1$ 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 $\pi_i$ 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 $\omega_i$ is a non-vanishing differential on $\mathrm{E}_i$. 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 $\mathrm{E}_i$, so that we may choose $\lambda_i$ above so that $\omega$ 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 $B_\infty$, 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 $\lambda$ is *not* a unit in $B_\infty$; however, the last term *is* well-defined and independent of the choices made because $\mathrm{E}_{p-1}$ is a lifting of the Hasse invariant, hence differs from $\lambda$ 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 $\omega$ by a unit in $B_\infty$. 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 $\mathrm{ord}(r) < p/(p+1)$, 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*

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

*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

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

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

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

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

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

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

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

Writing

$$\frac{rb_1}{E_{p-1}} = \frac{r_1^p b_1}{E_{p-1}} = r_1^{p-1}\frac{r_1 b_1}{E_{p-1}},$$

and noting that

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

we can write

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

where, since

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

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

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

Now, the $q$-expansion of $U(\frac{r_1^p b_1}{E_{p-1}})$ 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

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

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 $\mathrm{N} \geq 3$, $p \not| \mathrm{N}$, 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 $p \not| n$. In particular, if $f$ is an eigenform for the Hecke operators $\mathrm{T}_\ell$ 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* $\lambda$ 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 $f_\lambda$ clearly have the same $q$-expansion coefficients "outside of $p$". Thus, we have proved:

**Proposition II.3.14** *With the hypotheses above, for any $\lambda$ 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 $\mathrm{T}_\ell$ with $\ell \neq p$, then so is $f_\lambda$, 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 $\lambda$. 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* $\lambda$ 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 $\lambda$ 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 $\lambda$ 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 $(f_0, \lambda)$ for which $\lambda$ 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 $f_\lambda$ 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 $(f_0, \lambda)$ for which $\lambda \neq 0$ and the above construction produces the $q$-expansion of an overconvergent modular form, and the possible values of $\lambda$ form a sequence tending to zero. (It would be quite interesting to obtain an a priori criterion for determining whether a pair $(f_0, \lambda)$ 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 $\lambda$ is an eigenvalue of U, in which case $M(g) = M(\lambda)$ is the generalized eigenspace corresponding to the eigenvalue $\lambda$. 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 $\lambda$ is precisely the multiplicity of $\lambda^{-1}$ 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 $\psi_g$ 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 $\pi_g$ 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 $\pi_g$ and for $\psi_g$ are equivalent, and we will concentrate on the projection $\pi_g$.

Consider first the case where $g(X) = X - \lambda$. Let $h$ denote the multiplicity of $\lambda^{-1}$ as a root of $P(t)$, and let $\Delta$ 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 $\pi_g$ is the $h^{\text{th}}$ 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 $\mathrm{U}_1$. The assertion for $\psi_g$ 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 $\psi_g$ commute with the Hecke operators $\mathrm{T}_\ell$ 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 $\lambda$ 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 $\{\lambda_1, \lambda_2, \dots, \lambda_s\}$ 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 $\alpha$, which we call the "slope $\alpha$ eigenspace" of $\mathsf{M}(B,k,\mathrm{N};1) \otimes K$; we have obtained, for each $\alpha$, 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 $e_\alpha$; 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 $\alpha$ (respectively, less than $\alpha$), 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 $\mathrm{N}$ 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 $\alpha$ 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 $e_\alpha$ 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 $\lambda$ 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 $\lambda_i$ 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 $\operatorname{ord}(r) < 1/(p+1)$, 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

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

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 $\mathsf{k}$ 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 $\mathrm{E}_{p-1}$ 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:

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

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* $0 \le n \le d_1$, $\operatorname{ord}(c_n) \ge 0$,

*ii. if* $d_1 < n \le d_2$,

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

*iii. if* $d_2 < n \le d_3$,

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

*iv. if* $d_3 < n \le d_4$

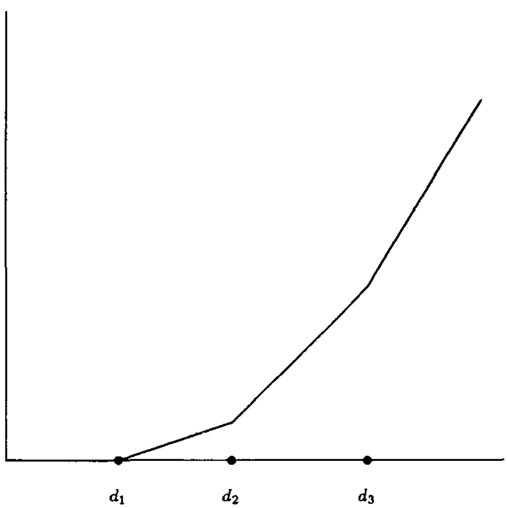
$$\operatorname{ord}(c_n) \ge \frac{m_2 + 3m_3 + 5(n - d_3)}{3} \ge \frac{m_2 + 3m_3}{3}$$

*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 $\alpha$? (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 $\alpha$, 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 $\mathrm{N}$ 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 $\alpha$ but not on $k$ or $\mathrm{N}$, 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 $\alpha$. 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 $e_\alpha$ 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 $e_\alpha$ on the weight $k$ and the parameter $r$. (Of course, $P(t)$ does not depend on $r$, but the norm of $e_\alpha$ 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 $e_\alpha$, 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 $\alpha$ 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 $\alpha$ 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 $\alpha$ 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 $\alpha$" 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 $\alpha$ 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 $e_\alpha$ to the slope $\alpha$ part, which we defined above. The question is whether the $e_\alpha$ "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 $\Lambda$-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 $\mathrm{N}$, 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 $\mathsf{T}_i$, 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.

# Chapter III

# Galois Representations

The focus of this chapter is the construction of Galois representations associated to $p$-adic modular forms. Specifically, we shall be interested the problem of studying the "deformations of a residual Galois representation" as considered by Mazur in [Ma]. Thus, we will begin with an absolutely irreducible Galois representation defined over a finite field, which we will assume to be attached, as in the work of Deligne and Serre, to a modular form defined over that field. We will then consider its liftings to complete noetherian overrings. In [Ma], Mazur constructed a universal lifting of this kind, and studied its properties in some detail. We will show that, under the assumption that the residual representation is attached to a modular form, a good portion of the liftings he obtains are in fact attached to $p$-adic modular forms.

The fact that one can attach $p$-adic representations to $p$-adic modular forms was first noticed by Hida, who, in [Hi86b] and [Hi86a], constructed analytic families of such representations attached to analytic families of ordinary $p$-adic modular forms; he also showed how to obtain a large number of such analytic families. This was further studied, still in the ordinary case, by Mazur and Wiles in [MW86], who constructed what may be called "the universal ordinary modular deformation", i.e., a family of deformations of a representation (which is assumed absolutely irreducible and attached to an ordinary modular form) which parametrizes all possible deformations attached to ordinary $p$-adic modular forms.

In this chapter, we continue in the spirit of these results, by constructing the universal modular deformation of the given (modular, absolutely irreducible) residual Galois representation, parametrizing all the deformations that are attached to $p$-adic modular forms of the given level. As in the case of Hida's work, the crucial step is to obtain a good theory of the duality between spaces of modular forms and their Hecke algebras, and we devote the first section of this chapter to constructing such a theory. We then outline a recent result of Hida which we feel should be better known, and proceed to consider the problem of constructing modular deformations of a given residual representation. Finally, we obtain an estimate on the dimension of the space of modular deformations, and consider its relation to the full space of deformations of the given residual representation. We conclude by formulating several questions that arise naturally from the theory as it is known at this point.

# III.1 Duality Theorems

Our first goal is to study the duality between spaces of modular forms and their Hecke algebras. Our goal is to identify $\mathbf{V}$ as a certain space of functions on the Hecke algebra $\mathbf{T}$, and similarly, $\mathbf{V}_{par}$ as a certain space of functions on the corresponding Hecke algebra $\mathbf{T}_0$. This works perfectly well in the case of parabolic modular functions, but is a little more complicated in the general case because of the presence of the constants in $\mathbf{V}$. The result is obtained by starting from V. Miller's results in the finite-dimensional case, and using the inverse limit technique of the section on Hecke operators. These results are similar to results obtained by Hida in [Hi86b], except that Hida's results give a Pontryagin duality while ours give a duality of topological $\mathbf{Z}_p$-modules. In particular, we show that there is a bijection between (normalized) eigenforms in $\mathbf{V}_{par}$ and continuous $\mathbf{Z}_p$-algebra homomorphisms from $\mathbf{T}_0$ to $p$-adic rings.

## III.1.1 Classical duality

To fix notation, let $\mathcal{H}_k(B, \mathrm{N}p^\nu)$ denote the Hecke algebra corresponding to the space $\mathsf{D}_k(B, \mathrm{N}p^\nu)$ of divided congruences of classical modular forms of weight at most $k$. We can define a bilinear form

$$\begin{array}{ccc} \mathcal{H}_k(B, \mathrm{N}p^\nu) \times \mathsf{D}_k(B, \mathrm{N}p^\nu) & \longrightarrow & B \\ (\mathrm{T}, f) & \longmapsto & a_1(\mathrm{T}f), \end{array} \tag{III.1}$$

where $a_1(\mathrm{T}f)$ denotes the coefficient of $q$ in the $q$-expansion of $\mathrm{T}f$. If we let $\mathsf{S}^k(B, \mathrm{N}p^\nu) \subset \mathsf{D}_k(B, \mathrm{N}p^\nu)$ denote the subspace of divided congruences of cusp forms , and let $\mathrm{h}_k(B, \mathrm{N}p^\nu)$ denote the corresponding Hecke algebra, then (III.1) clearly induces a bilinear form

$$\mathrm{h}_k(B, \mathrm{N}p^\nu) \times \mathsf{S}^k(B, \mathrm{N}p^\nu) \longrightarrow B.$$

Similarly, we get a bilinear form

$$\mathcal{H}'_k(B, \mathrm{N}p^\nu) \times \mathsf{D}'_k(B, \mathrm{N}p^\nu) \longrightarrow B,$$

where the primes have the same meaning as above, i.e., if $K$ is the fraction field of $B$, then

$$\mathsf{D}'_k(K, \mathrm{N}p^\nu) = \bigoplus_{i=1}^{k} M(K, i, \mathrm{N}p^\nu),$$

$$\mathsf{D}'_k(B, \mathrm{N}p^\nu) = \{f \in \mathsf{D}'_k(K, \mathrm{N}p^\nu) | f(q) \in \mathbf{Z}_p[[q]], \}$$

and $\mathcal{H}'_k$ is the corresponding Hecke algebra. Then we have the following result:

**Theorem III.1.1** *Let $K$ be a finite extension of $\mathbf{Q}_p$, and let $\mathcal{O}_K$ be its ring of integers. Assume $B$ is either $K$ or $\mathcal{O}_K$, and let*

$$m_k(\mathcal{O}_K, \mathrm{N}p^\nu) = (K + \mathsf{D}_k(\mathcal{O}_K, \mathrm{N}p^\nu))/K.$$

*Then the pairing (III.1) induces perfect pairings of $B$-modules*

$$\begin{array}{rcl} \mathcal{H}'_k(K, \mathrm{N}p^\nu) \times \mathsf{D}'_k(K, \mathrm{N}p^\nu) & \longrightarrow & K \\ \mathcal{H}'_k(\mathcal{O}_K, \mathrm{N}p^\nu) \times m_k(\mathcal{O}_K, \mathrm{N}p^\nu) & \longrightarrow & \mathcal{O}_K \\ \mathrm{h}_k(B, \mathrm{N}p^\nu) \times \mathsf{S}^k(B, \mathrm{N}p^\nu) & \longrightarrow & B. \end{array}$$

*Proof*: This is [Hi86b, Proposition 2.1], where the result is attributed to V. Miller. □

One should note that Hida shows also that $\mathcal{H}'_k(\mathcal{O}_K, \mathrm{N}p^\nu)$ (which is defined as the Hecke algebra corresponding to $\mathsf{D}'_k(\mathcal{O}_K, \mathrm{N}p^\nu)$) "is" the Hecke algebra corresponding to $m_k(\mathcal{O}_K, \mathrm{N}p^\nu)$ (the subalgebra of the ring of endomorphisms generated by the Hecke and diamond operators); this follows from the (obvious) fact that

$$m_k(\mathcal{O}_K, \mathrm{N}p^\nu) \otimes K = \mathsf{D}'_k(\mathcal{O}_K, \mathrm{N}p^\nu) \otimes K = \mathsf{D}'_k(K, \mathrm{N}p^\nu) = \bigoplus_{i=1}^{k} M(K, i, \mathrm{N}p^\nu).$$

In particular, we have, setting $\mathsf{S}^k = \mathsf{S}^k(\mathbf{Z}_p, \mathrm{N}p^\nu)$ and $\mathrm{h}_k = \mathrm{h}_k(\mathbf{Z}_p, \mathrm{N}p^\nu)$, we get

$$\mathsf{S}^k \cong \mathrm{Hom}_{\mathbf{Z}_p}(\mathrm{h}_k, \mathbf{Z}_p)$$

(homomorphisms of $\mathbf{Z}_p$-modules), and it is easy to see that a cusp form $f \in \mathsf{S}^k$ corresponds to a homomorphism *of $\mathbf{Z}_p$-algebras* if and only if it is a normalized eigenform, that is, if and only if it is a simultaneous eigenform for all the Hecke and diamond operators and has $a_1(f) = 1$. Conversely, every algebra homomorphism $\mathrm{h}_k \longrightarrow \mathbf{Z}_p$ corresponds to a normalized eigenform (so that given a system of eigenvalues in $\mathbf{Z}_p$, there is a unique normalized eigenform belonging to it).

We would like to extend this to other $p$-adic rings B. For this, note that, since $\mathrm{h}_k$ is $\mathbf{Z}_p$-free,

$$\mathsf{S}^k \otimes B \cong \mathrm{Hom}_{\mathbf{Z}_p}(\mathrm{h}_k, \mathbf{Z}_p) \otimes B \cong \mathrm{Hom}_{\mathbf{Z}_p}(\mathrm{h}_k, B).$$

We will use this fact later, when we pass to the limit situation.

### III.1.2 Duality for parabolic $p$-adic modular functions

Given the pairing on the finite-dimensional case, one may attempt to pass to the limit in order to obtain results for generalized $p$-adic modular functions. We consider first the case of parabolic modular functions, which is clearly simpler (because, as we saw above, the constants are a problem: for any constant $b \in B$ and any $\mathrm{T} \in \mathbf{T}$, we have $(\mathrm{T}, b) \mapsto 0$ under the pairing III.1)

Assume first that $B = \mathcal{O}_K$ as above, so that we have a perfect pairing

$$\mathfrak{h}_k(B, \mathrm{N}p^\nu) \times \mathsf{S}^k(B, \mathrm{N}p^\nu) \longrightarrow B.$$

The $\mathfrak{h}_k$ form an inverse system of $\mathbf{Z}_p$-algebras, while the $\mathsf{S}^k$ form a direct system of $\mathbf{Z}_p$-modules. Going to the limit, we get a pairing

$$\varprojlim_k \mathfrak{h}_k(B, \mathrm{N}p^\nu) \times \varinjlim_k \mathsf{S}^k(B, \mathrm{N}p^\nu) = \mathbf{T}_0(B, \mathrm{N}) \times \mathsf{S}(B, \mathrm{N}p^\nu) \longrightarrow B,$$

which gives maps

$$\mathbf{T}_0(B, \mathrm{N}) \xrightarrow{\sim} \mathrm{Hom}_B(\mathsf{S}(B, \mathrm{N}p^\nu), B) \xrightarrow{\sim} \mathrm{Hom}_B(\mathbf{V}_{par}(B, \mathrm{N}p^\nu), B),$$

and

$$\mathsf{S}(B, \mathrm{N}p^\nu) \longrightarrow \mathrm{Hom}_B(\mathbf{T}_0(B, \mathrm{N}), B).$$

It is clear that the first map, being the inverse limit of the maps

$$\mathfrak{h}_k(B, \mathrm{N}p^\nu) \xrightarrow{\sim} \mathrm{Hom}_B(\mathsf{S}^k(B, \mathrm{N}p^\nu), B),$$

is, as indicated, an isomorphism; the second isomorphism is immediate since $\mathsf{S}$ is $p$-adically dense in $\mathbf{V}_{par}$. As to the map

$$\mathsf{S} \longrightarrow \mathrm{Hom}_B(\mathbf{T}_0, B),$$

we know only that it is injective, since it is the direct limit of the maps

$$\mathsf{S}^k(B, \mathrm{N}p^\nu) \xrightarrow{\sim} \mathrm{Hom}_B(\mathfrak{h}_k(B, \mathrm{N}p^\nu), B).$$

It therefore identifies $\mathsf{S}$ with a submodule of $\mathrm{Hom}_B(\mathbf{T}_0(B, \mathrm{N}p^\nu), B)$, which is easily seen to be the submodule of all the $B$-module homomorphisms $f : \mathbf{T}_0 \longrightarrow B$ which factor through the projection $\mathbf{T}_0 \longrightarrow \mathfrak{h}_k$ for some $k$ (since $\mathsf{S}$ is just the union of the $\mathsf{S}^k$); any such homomorphism will be continuous if we give $\mathbf{T}_0$ its inverse limit topology (which makes it compact) and $B$ its $p$-adic topology. Let us denote this submodule by $\mathrm{Hom}_B^{fact}(\mathbf{T}_0, B)$; it clearly depends on the particular representation of $\mathbf{T}_0$ as an inverse limit (as does $\mathsf{S}$). The $p$-adic topology on $\mathsf{S}$ corresponds to the topology of uniform convergence on $\mathrm{Hom}_B^{fact}(\mathbf{T}_0, B)$ (i.e., to the sup norm induced by the $p$-adic norm on B). Since $\mathbf{V}_{par}$ is the $p$-adic completion of $\mathsf{S}$, taking completions induces an identification between $\mathbf{V}_{par}$ and the completion of $\mathrm{Hom}_B^{fact}(\mathbf{T}_0, B)$; this last is contained in the submodule $\mathrm{Hom}_{B,conts}(\mathbf{T}_0, B)$ of *continuous* $B$-module homomorphisms (where $\mathbf{T}_0$ is given the inverse limit topology and $B$ the $p$-adic topology), which is complete (with the topology of uniform convergence). Thus, we have obtained an inclusion $\mathbf{V}_{par} \hookrightarrow \mathrm{Hom}_{B,conts}(\mathbf{T}_0, B)$, mapping a parabolic modular function $f$ to the homomorphism $\phi_f$ defined by $\phi_f(\mathrm{T}) = a_1(\mathrm{T}f)$.

**Proposition III.1.2** *Let $B = \mathcal{O}_K$ be the ring of integers in a finite extension $K$ of $\mathbf{Q}_p$, and let $\mathbf{T}_0(B,\mathrm{N})$ have the inverse limit topology, $B$ and $\mathbf{V}_{par}(B,\mathrm{N})$ the $p$-adic topology, and $\mathrm{Hom}_{B,conts}(\mathbf{T}_0(B,\mathrm{N}),B)$ the topology of uniform convergence. Then the mapping*

$$\mathbf{V}_{par}(B,\mathrm{N}) \longrightarrow \mathrm{Hom}_{B,conts}(\mathbf{T}_0(B,\mathrm{N}),B)$$

*defined by $f \mapsto \phi_f$ is an isomorphism of topological $B$-modules.*

*Proof*: It suffices, after the above discussion, to show that any continuous homomorphism $\phi : \mathbf{T}_0 \longrightarrow B$ can be approximated by homomorphisms which factor through one of the $\mathrm{h}_k$. Given $\phi$, consider its reduction mod $p^n$,

$$\phi_n : \mathbf{T}_0 \longrightarrow B/p^nB.$$

Since $B/p^nB$ is (finite and) discrete, it is clear that $\phi_n$ factors through some $\mathrm{h}_k$, giving a map $\overline{\psi}_n : \mathrm{h}_k \to B/p^nB$, which can then be lifted to a map $\psi_n : \mathrm{h}_k \longrightarrow B$, because $\mathrm{h}_k$ is a free $B$-module. Then it is clear that $\psi_n \to \phi$, and we are done. □

The restriction to the case when $B = \mathcal{O}_K$ in the above result can be removed without too much trouble by using the fact that, for *any* $p$-adic ring $B$, $\mathbf{V}_{par}(B,\mathrm{N}) = \mathbf{V}_{par}(\mathbf{Z}_p,\mathrm{N})\hat{\otimes}B$, which reduces everything to the case of $\mathbf{Z}_p$. More generally, we can restrict to algebras over the Witt ring $W(\mathsf{k})$ of a finite field $\mathsf{k}$, and get a general duality statement. Recall that we say an parabolic eigenform for the Hecke algebra $\mathbf{T}_0$ is *normalized* if the coefficient $a_1$ of $q$ in its $q$-expansion is equal to 1. Then we have:

**Corollary III.1.3** *Let $\mathsf{k}$ be a finite field, and let $W(\mathsf{k})$ be its ring of Witt vectors. For any $p$-adically complete $W(\mathsf{k})$-algebra $B$ (with the $p$-adic topology), we have*

$$\mathbf{V}_{par}(B,\mathrm{N}) \cong \mathrm{Hom}_{W(\mathsf{k}),conts}(\mathbf{T}_0(W(\mathsf{k}),\mathrm{N}),B)$$

*(continuous homomorphisms of $W(\mathsf{k})$-modules), via the map $f \mapsto \phi_f$. Moreover, $\phi_f$ is a homomorphism of $W(\mathsf{k})$-*algebras *if and only if $f \in \mathbf{V}_{par}(B,\mathrm{N})$ is a normalized simultaneous eigenform for the Hecke and diamond operators. In particular, given any eigenform for $\mathbf{T}_0$, there exists a* normalized *eigenform with the same system of eigenvalues.*

*Proof*: For the first assertion,

$$\begin{aligned}
\mathbf{V}_{par}(B,\mathrm{N}) &= \varprojlim_n \mathbf{V}_{par}(W(\mathsf{k}),\mathrm{N}) \otimes B/p^nB \\
&= \varprojlim_n \mathsf{S}(W(\mathsf{k}),\mathrm{N}) \otimes B/p^nB \\
&\cong \varprojlim_n \mathrm{Hom}_{W(\mathsf{k}),conts}(\mathbf{T}_0(W(\mathsf{k}),\mathrm{N}),\mathbf{Z}_p) \otimes B/p^nB \\
&= \varprojlim_n \mathrm{Hom}^{fact}_{W(\mathsf{k})}(\mathbf{T}_0(W(\mathsf{k}),\mathrm{N}),\mathbf{Z}_p) \otimes B/p^nB \\
&= \varprojlim_n \mathrm{Hom}^{fact}_{W(\mathsf{k})}(\mathbf{T}_0(W(\mathsf{k}),\mathrm{N}),B/p^nB) \\
&= \varprojlim_n \mathrm{Hom}_{W(\mathsf{k}),conts}(\mathbf{T}_0(W(\mathsf{k}),\mathrm{N}),B/p^nB) \\
&= \mathrm{Hom}_{W(\mathsf{k}),conts}(\mathbf{T}_0(W(\mathsf{k}),\mathrm{N}),B).
\end{aligned}$$

The second assertion is immediate from the definition of $\phi_f$, and the third is clear, since any eigenform defines an algebra homomorphism

$$\mathbf{T}_0(W(\mathsf{k}),\mathrm{N}) \longrightarrow B$$

(its system of eigenvalues). □

In most situations, it will be sufficient with the case $W(\mathsf{k}) = \mathbf{Z}_p$, which is the most general. In the theory of deformations of Galois representations, however, we will want to base-change to $W(\mathsf{k})$.

In particular, the Corollary applies to classical cuspforms. Hence, for *any* $k$ and $\nu$, a classical cuspform $f$ of level $\mathrm{N}p^\nu$, weight $k$, and defined over B, determines a continuous homomorphism $\phi_f$ of $\mathbf{Z}_p$-modules $\mathbf{T}_0 \longrightarrow B$. If $f$ is a normalized eigenform for $\mathbf{T}_0$ (hence in particular has a nebentypus), then $\phi_f$ is an algebra homomorphism whose restriction to $G(\mathrm{N})$ is the character determined by the weight and nebentypus[1].

**Remark:** The reference to the various topologies in the statement of this general duality result should be carefully noted. A parabolic generalized $p$-adic modular function defined over a $p$-adic ring $B$ may be identified with a continuous $\mathbf{Z}_p$-module homomorphism $\mathbf{T}_0 \longrightarrow B$, provided one gives $\mathbf{T}_0$ its inverse limit topology (of which we give a more intrinsic description below) and $B$ its $p$-adic topology. For example, the identity map $\mathbf{T}_0 \xrightarrow{id} \mathbf{T}_0$ does *not* correspond to a modular function (because the $p$-adic topology is strictly finer than the inverse limit topology); in other words, there is no "universal" parabolic modular function. This leads to the definition of a family of modular functions in the next section.

Since the inverse limit topology on $\mathbf{T}_0$ appears in such a central manner in the discussion above, one would like to be able to give an intrinsic characterization of it, independent of the particular description of $\mathbf{T}_0$ as an inverse limit of finite $\mathbf{Z}_p$-algebras (of which there are many, since one may work with level $\mathrm{N}p^\nu$ for any $\nu \geq 1$, and see also section III.3). This is indeed possible, at least when $B = \mathcal{O}_K$ for some finite extension $K/\mathbf{Q}_p$.

**Proposition III.1.4** *Let $B = \mathcal{O}_K$ for some finite extension $K/\mathbf{Q}_p$. Under the isomorphism*

$$\mathbf{T}_0(B,\mathrm{N}) \xrightarrow{\sim} \mathrm{Hom}_{\mathbf{Z}_p}(\mathbf{V}_{par}, B),$$

*the inverse limit topology of $\mathbf{T}_0(B,\mathrm{N})$ corresponds to the compact-open topology on $\mathrm{Hom}_{\mathbf{Z}_p}(\mathbf{V}_{par}, B)$ (where $\mathbf{V}_{par}$ and $B$ are given their $p$-adic topologies).*

[1] In the preceding chapter, we recalled Hida's theory of the *ordinary* part $e\mathbf{T}_0$ of the Hecke algebra, which is a direct summand of $\mathbf{T}_0$, and corresponds to the "unit-root eigenspace" for the U operator. It is immediate, then, that if the eigenform $f$ satisfies $\mathrm{U}f = \lambda f$ with $\lambda$ a $p$-adic unit, then $\phi_f$ factors through the ordinary algebra $e\mathbf{T}_0$.

*Proof*: The inverse limit topology is generated by open sets of the form

$$\{\mathrm{T} \in \mathbf{T}_0 \,|\, \mathrm{T}f \in \mathcal{U},\ \forall f \in \mathsf{S}^k\},$$

where $\mathcal{U}$ runs over the open subsets of $B$ and $k$ runs over the integers, which are open in the compact-open topology (because the $\mathsf{S}^k$, being free of finite rank over $\mathbf{Z}_p$, are compact in the $p$-adic topology). Hence, it is clear that the compact-open topology is finer than the inverse limit topology.

For the converse, we want to show that sets of the form

$$\{\mathrm{T} \in \mathbf{T}_0 \,|\, \mathrm{T}(K) \subset \mathcal{U}\},$$

($K \subset \mathbf{V}_{par}$ compact, $\mathcal{U} \subset B$ open) are open in the inverse limit topology. It is clearly enough to consider the case where $K$ is a compact submodule of $\mathbf{V}_{par}$ and $\mathcal{U} = p^n B$. Since $K$ is compact in the $p$-adic topology and $\mathbf{Z}_p$-free, it must be $\mathbf{Z}_p$-free of finite rank; let $(f_1, \ldots, f_r)$ be a basis. Since $\mathsf{S}$ is dense in $\mathbf{V}_{par}$, one may choose $(g_1, \ldots, g_r)$ such that $g_i \in \mathsf{S}^k$ (for some fixed, sufficiently large $k$) and $g_i \equiv f_i \pmod{p^\nu}$ with $\nu > n$. Then it is clear that

$$\begin{aligned}\{\mathrm{T} \,|\, \mathrm{T}(K) \subset p^n B\} &= \{\mathrm{T} \,|\, \mathrm{T}f_i \in p^n B,\ i = 1, \ldots, r\} \\ &= \{\mathrm{T} \,|\, \mathrm{T}g_i \in p^n B,\ i = 1, \ldots, r\},\end{aligned}$$

which is clearly open in the inverse limit topology (it is the inverse image under the canonical projection of an open subset of $\mathfrak{h}_k$). Thus, the two topologies are equal. □

## III.1.3 The non-parabolic case

To get analogous results for the full ring of $p$-adic modular functions (not necessarily parabolic), one must somehow get around the fact that the constants in $\mathbf{V}$ will necessarily pair to zero in the pairing (III.1). This turns out not to be too difficult, and since we will not use it later, we will only sketch the results.

Let $K$ be a finite extension of $\mathbf{Q}_p$, and let $B = \mathcal{O}_K$. We have already seen that we have the perfect pairing

$$\mathcal{H}'_k(B, \mathrm{N}) \times m_k(B, \mathrm{N}) \longrightarrow B,$$

where $m_k$ is defined as above, rather than a pairing for $\mathsf{D}'_k$, since it is clear that a sum of modular forms over K (the field of fractions of B) that has integral $q$-expansion except for its $a_0$ term will define a map $\mathcal{H}'_k \longrightarrow B$. Note that

$$m_k(B, \mathrm{N}) \cong \{f \in \mathsf{D}'_k(K, \mathrm{N}) | f(q) \in K + qB[[q]]\},$$

so that going to $m_k(B, \mathrm{N})$ gives precisely the desired space.

It is again possible to pass to the limit situation. Since we have the obvious maps $m_k(B,\mathrm{N}) \hookrightarrow m_{k+1}(B,\mathrm{N})$, we may consider the limit, and define

$$m_\infty(B,\mathrm{N}) = \varinjlim_k m_k(B,\mathrm{N}).$$

Then note that

$$m_\infty(B,\mathrm{N}) = \varinjlim_k \frac{\mathsf{D}_k(B,\mathrm{N})+K}{K} = \frac{\mathsf{D}(B,\mathrm{N})+K}{K},$$

so that $m_\infty(B,\mathrm{N})$ is not a completely mysterious space. (It may also be described as in the preceding paragraph, as the space of sums of modular forms over $K$ which have integral $q$-expansion except possibly for the first coefficient.)

As before, we get a pairing

$$\mathbf{T}(B,\mathrm{N}) \times m_\infty(B,\mathrm{N}) \longrightarrow B,$$

which, as before, gives a duality between $\mathbf{T}(B,\mathrm{N})$ with its inverse image topology and $\overline{m}_\infty(B,\mathrm{N})$, where the bar indicates the $p$-adic completion. (We have here used the fact that

$$\mathsf{D}' = \varinjlim_k \mathsf{D}'_k$$

is dense in $\mathbf{V}$, and hence that its Hecke algebra is the full Hecke algebra, as noted above.) The only difficulty, then, is to determine the relation between $\mathbf{V}$ and the $p$-adic completion of $m_\infty$. The elements of $\overline{m_\infty}$ are limits of sequences $f_i \in m_\infty(B,\mathrm{N})$ of modular forms whose $q$-expansions are integral except perhaps for the $a_0$ term, and the $p$-adic norm is the $p$-adic norm on $q$-expansions shorn of their $a_0$ term, i.e., if $f = \sum a_n q^n$, with $a_n$ $p$-integral for $n \geq 1$, then we set

$$\|f\| = \sup_{n\geq 1} |a_n|,$$

where $|\cdot|$ denotes the $p$-adic norm on B. Giving $m_\infty(B,\mathrm{N})$ this topology, we have a continuous surjective Hecke-equivariant map

$$\mathsf{D}(B,\mathrm{N})+K \longrightarrow m_\infty(B,\mathrm{N}),$$

which extends by continuity to a map

$$\mathbf{V}(B,\mathrm{N})+K \longrightarrow \overline{m}_\infty(B,\mathrm{N}).$$

We claim this is still onto. Let $[f]$ denote the image in $m_\infty$ of $f \in \mathsf{D}+K$; it is clear that $\|[f]\| \leq \|f\|$, where of course we use the correct norm in each space. However, one can make a canonical choice of the lifting $f$ by requiring $f(q) \in qB[[q]]$ (which, since the constants belong to $K$, can always be done!), in which case we will have $\|f\| = \|[f]\|$. Note also that such a lifting $f$ will belong to $\mathsf{D}$, and not only to $\mathsf{D}+K$. Thus, given a Cauchy sequence in $m_\infty$, we can lift it to a Cauchy sequence in $\mathsf{D}$, whose limit will be mapped to the limit of the given sequence, so that the map on completions is onto. The kernel consists precisely of the constants $K \subset \mathbf{V} + K$, so that we get an exact sequence of Hecke-equivariant $\mathbf{Z}_p$-algebra homomorphisms

$$0 \longrightarrow K \longrightarrow \mathbf{V}(B,\mathrm{N}) + K \longrightarrow \overline{m}_\infty(B,\mathrm{N}) \longrightarrow 0.$$

To summarize, we have shown:

**Proposition III.1.5** *Let $K$ be a finite extension of $\mathbf{Q}_p$, and let $B = \mathcal{O}_K$. Then the pairing*

$$\begin{array}{ccc} \mathbf{V}(B,\mathrm{N}) \times \mathbf{T}(B,\mathrm{N}) & \longrightarrow & B \\ (f,\mathrm{T}) & \mapsto & a_1(\mathrm{T}f) \end{array}$$

*induces isomorphisms*

$$\mathbf{T}(B,\mathrm{N}) \cong \mathrm{Hom}_B(\mathbf{V}_1(B,\mathrm{N}), B)$$

*(continuous homomorphisms of $B$-modules) and*

$$\mathbf{V}_1(B,\mathrm{N}) \cong \mathrm{Hom}_{B,conts}(\mathbf{T}(B,\mathrm{N}), B),$$

*where we give $\mathbf{T}(B,\mathrm{N})$ the inverse limit topology, and where*

$$\mathbf{V}_1(B,\mathrm{N}) = \frac{\mathbf{V}(B,\mathrm{N}) + K}{K}.$$

Thus a continuous $B$-module homomorphism $\mathbf{T}(B,\mathrm{N}) \longrightarrow B$ determines a unique element of $\mathbf{V}_1(B,\mathrm{N})$; however, in general there is no canonical Hecke-equivariant way to determine a lifting to $\mathbf{V}(B,\mathrm{N})$. If the map is a $B$-*algebra* homomorphism, the corresponding element of $\mathbf{V}_1(B,\mathrm{N})$ is an eigenform (up to constants), and hence has a well-defined weight. Provided the weight is not zero (i.e., provided $G(\mathrm{N})$ is not mapped to 1), we can then choose canonically a constant term in the $q$-expansion, so that such a map corresponds to a well-defined eigenform in $\mathbf{V}(B,\mathrm{N})[1/p]$ which has $q$-expansion $f(q) \in K + qB[[q]]$. The example of

$$f = \frac{1}{p}\mathrm{E}_{p-1}$$

shows that it is not possible, in general, to assume that the resulting eigenform is in $\mathbf{V}(B,\mathrm{N})$ itself.

Applying the preceding proposition to the case $B = \mathbf{Z}_p$ and using an inverse limit argument analogous to that in the proof of Corollary III.1.3, we get an analogous result. Recall that $\mathbf{T} = \mathbf{T}(\mathbf{Z}_p,\mathrm{N})$, and give $\mathbf{T}$ its inverse limit topology; then:

**Corollary III.1.6** *For any $p$-adic ring $B$ with the $p$-adic topology, we have*

$$\begin{array}{c} \mathbf{V}_1(B,\mathrm{N}) \cong \mathrm{Hom}_{\mathbf{Z}_p,conts}(\mathbf{T}, B) \\ f \mapsto \phi_f \end{array}$$

*(continuous homomorphisms of $\mathbf{Z}_p$-modules), where*

$$\phi_f(\mathrm{T}) = a_1(\mathrm{T}f)$$

*and*

$$\mathbf{V}_1(B,\mathrm{N}) = \frac{\mathbf{V}(B,\mathrm{N}) + K}{K}.$$

*Moreover, $\phi_f$ is an algebra homomorphism if and only if $f \in \mathbf{V}_1(B,\mathrm{N})$ is a normalized simultaneous eigenform for the Hecke and diamond operators (up to constants). In this case, $f \in \mathbf{V}_1(B,\mathrm{N})$ may be canonically lifted to an eigenform $f \in \mathbf{V}(B,\mathrm{N})[1/p]$ whose $q$-expansion satisfies $f(q) \in K + qB[[q]]$, provided $\phi_f(G(\mathrm{N})) \neq 1$, i.e., provided $f$ is not of weight zero.*

In what follows, we will usually prefer to consider only the case of parabolic modular functions, since the duality theory is then much simpler, but will occasionally mention the general case.

# III.2 Families of Modular Forms

As we have observed in the preceding section, the general duality between parabolic $p$-adic modular functions and the Hecke algebra $\mathbf{T}_0$ is complicated by the necessity of keeping the topologies involved straight. The point of this section is to introduce a concept that alleviates the problem somewhat. As before, everything makes sense (and is true) for $p$-adically complete algebras over the Witt ring of a finite field; we will consider only the case $\mathsf{k} = \mathbf{F}_p$ (so that $W(\mathsf{k}) = \mathbf{Z}_p$), and leave the extension to the reader.

The concept of an "analytic family of $p$-adic modular forms" was first introduced by Serre in [Se73]. He considered there the Iwasawa algebra $\Lambda = \mathbf{Z}_p[[\Gamma]]$ (the completed group ring with coefficients in $\mathbf{Z}_p$ of the pro-$p$-group $\Gamma = 1 + p\mathbf{Z}_p \subset \mathbf{Z}_p^\times$), given the inverse limit topology, with which it is a compact $\mathbf{Z}_p$-algebra. Recall that choosing a topological generator $\gamma$ of $\Gamma$ defines an isomorphism of topological $\mathbf{Z}_p$-algebras

$$\begin{array}{ccc} \mathbf{Z}_p[[\Gamma]] & \longrightarrow & \mathbf{Z}_p[[T]] \\ \langle\gamma\rangle & \longmapsto & 1+T, \end{array}$$

where $\mathbf{Z}_p[[T]]$ is given the $(p,T)$-adic topology, and where we use angular brackets to distinguish elements of the group $\Gamma$ from elements of $\mathbf{Z}_p$ (so that $\langle\gamma\rangle - \gamma$ is a nonzero element of the group ring). Then Serre defined an analytic family of modular forms to be a formal $q$-expansion

$$F(q) = A_0 + A_1 q + A_2 q^2 + \dots$$

where $A_j \in \Lambda$, such that, for every $k \in \mathbf{Z}_p$ and for a fixed $i \in \mathbf{Z}/(p-1)\mathbf{Z}$, the image of $F(q)$ under the map $\Lambda \longrightarrow \mathbf{Z}_p$ defined by $\langle\gamma\rangle \mapsto \gamma^k$ is a $p$-adic modular form of weight $(i,k)$, i.e., belongs to $\mathsf{M}(\mathbf{Z}_p, \chi_{(i,k)}, \mathrm{N}; 1)$. To get a family of cusp forms, of course, we would require $A_0 = 0$ and that each specialization be a cusp form.

To translate this definition to our situation (in the cuspidal case), note that $\mathbf{T}_0$ is naturally an algebra over the algebra $\mathbf{\Lambda} = \mathbf{Z}_p[[\mathbf{Z}_p^\times]]$ topologically generated by the diamond operators $\langle x, 1\rangle$ for $x \in \mathbf{Z}_p$. Note that $\mathbf{\Lambda} = \Lambda[\mathbf{Z}/(p-1)\mathbf{Z}]$ is just a group ring over $\Lambda$, which (because $p \not| (p-1)$) we may decompose according to the powers of the Teichmüller character (which are the characters of $\mathbf{Z}/(p-1)\mathbf{Z}$); we write

$$\mathbf{\Lambda} = \bigoplus_{i \bmod (p-1)} \Lambda_{(i)}.$$

Then it is clear that an analytic family of $p$-adic modular forms in the sense above is a map $\mathbf{T}_0 \longrightarrow \Lambda_{(i)}$ which, when composed with the canonical maps $\Lambda \longrightarrow \mathbf{Z}_p$ given by $\langle\gamma\rangle \mapsto \gamma^k$, gives continuous maps $\mathbf{T}_0 \longrightarrow \mathbf{Z}_p$. This amounts to a map $\mathbf{T}_0 \longrightarrow \Lambda_{(i)}$ which is continuous when we give $\Lambda$ its *inverse limit* topology. Thus, an analytic family of $p$-adic modular forms is *not* necessarily a generalized $p$-adic modular function defined over $\Lambda$, though it defines $p$-adic modular forms by "specialization to weight $k$" for every $k$. (The distinction may be understood as follows: a generalized $p$-adic modular function defined over $\Lambda$ can be evaluated at any trivialized elliptic curve defined over *any* $p$-adically complete $\Lambda$-algebra; a map $\mathbf{T}_0 \longrightarrow \Lambda$ does not determine such a rule. Consider, for example, a trivialized curve defined over $\Lambda \otimes \mathbf{F}_p \cong \mathbf{F}_p[[T]]$.)

Given this re-interpretation of Serre's definition, it is clear that we may extend it as follows.

**Definition III.2.1** *Let* $\mathbf{B}$ *be a* $p$*-adically complete topological* $\mathbf{Z}_p$*-algebra. A* $\mathbf{B}$*-valued family of parabolic* $p$*-adic modular functions is a continuous* $\mathbf{Z}_p$*-module homomorphism* $\mathbf{f} : \mathbf{T}_0 \longrightarrow \mathbf{B}$*. Given any continuous map* $\phi : \mathbf{B} \longrightarrow B$ *to a* $p$*-adic ring* $B$ *(where we give* $B$ *the* $p$*-adic topology), we denote by* $\mathbf{f}_\phi$ *the modular function defined over* $B$ *corresponding to the composite homomorphism* $\phi \circ \mathbf{f}$*; we call* $\mathbf{f}_\phi$ *the "specialization via* $\phi$*" (sometimes "to weight* $\phi$*") of the family* $\mathbf{f}$*. Finally, we say* $\mathbf{f}$ *is a family of eigenforms if every specialization* $\mathbf{f}_\phi$ *is a simultaneous eigenfunction for the Hecke, diamond, and* U *operators (and hence for the action of* $\mathbf{T}_0$*).*

**Remarks:**

i. Of course, if the map $\mathbf{f}$ is continuous when we give $\mathbf{B}$ the $p$-adic topology, then we simply obtain a parabolic modular function defined over $\mathbf{B}$. This makes $\mathbf{B}$-valued families of modular functions a generalization of modular functions defined over $\mathbf{B}$.

ii. In many cases, the family $\mathbf{f}$ is determined by the set of all its specializations (for example, when $\mathbf{B}=\mathbf{\Lambda}$); when that is true, we will sometimes confuse the family $\mathbf{f}$ with the set $\{\mathbf{f}_\chi : \chi \in \mathrm{Hom}_{conts}(\mathbf{Z}_p^\times, \mathbf{Z}_p^\times)\}$ of its specializations.

iii. A family of parabolic modular functions has a $q$-expansion; simply define

$$\mathbf{f}(q) = \sum_{n\geq 1} \mathbf{f}(\mathrm{T}_n) q^n.$$

iv. The "universal family" of parabolic modular functions is simply the identity map $\mathbf{T}_0 \longrightarrow \mathbf{T}_0$.

The most interesting case of the above will be when $\mathbf{B}$ is a $\mathbf{\Lambda}$-algebra, in which case one can consider, for each $\chi \in \mathrm{Hom}_{conts}(\mathbf{Z}_p^\times, \mathbf{Z}_p^\times)$, the map

$$\phi_\chi : \mathbf{B} \longrightarrow B_\chi = \mathbf{B}/P_\chi\mathbf{B},$$

where $P_\chi\mathbf{B}$ denotes the ideal generated by the elements $\langle x, 1\rangle - \chi(x) \in \mathbf{\Lambda}$. If these maps are continuous when we give the quotient the $p$-adic topology, we call them "specialization to weight $\chi$", and write $\mathbf{f}_\chi$ for the corresponding specializations of a family of modular forms $\mathbf{f}$. Note that we then have $\mathbf{f}_\chi \in \mathsf{M}(B_\chi, \chi, \mathrm{N}; 1)$. For example, if $\mathbf{B} = \mathbf{\Lambda}$, we have $B_\chi = \mathbf{Z}_p$ for all $\chi$, and we recover Serre's situation.

**Remark:** There is one important aspect in which the above definition is weaker than Serre's: the restriction to parabolic modular functions. In fact, the best-known example of an analytic family, Serre's Eisenstein family $G_\chi$, is not parabolic. The problem, of course, is the fact that the duality between modular functions and Hecke operators gets complicated in the non-parabolic case. In particular, it is not clear that a family of eigenforms $\mathbf{f} : \mathbf{T} \longrightarrow \Lambda$ in our sense determines a family of eigenforms in Serre's sense, i.e., that there is a well-defined $A_0 \in \Lambda$ which is the constant term in the $q$-expansion of $\mathbf{f}$. In fact, this is not always true; given the duality theorems, the most we can expect is an element of the fraction field of $\Lambda$. Serre has shown (in [Se73]) that we have two cases. First, suppose $i$ is not divisible by $p-1$. Then, given modular functions $F_k$ of $p$-adic weight $(i, k)$, they are the specializations of an analytic family of modular forms ("with values in $\Lambda$") if and only if there are elements $A_i \in \Lambda$, $i \geq 1$, such that $A_i$ specializes (via $\gamma \mapsto \gamma^k$) to the $i^{\text{th}}$ coefficient of the $q$-expansion of $F_k$ for each $k$. (In other words, if the coefficients $a_1(F_k), a_2(F_k), \ldots$ are specializations of elements of $\Lambda$, then so is the first coefficient $a_0(F_k)$.) This avoids the difficulties with the integrality of the zero-th coefficient, and shows that in this case an analytic family of (non-parabolic) $p$-adic modular forms with values in $\Lambda$ *is* a continuous map $\mathbf{T} \longrightarrow \Lambda$. In the case when $i \equiv 0 \pmod{p-1}$, however, the Eisenstein family already shows that this cannot be the case; in this case Serre shows that if $a_1(F_k), a_2(F_k), \ldots$ are specializations of elements of $\Lambda$, then $a_0(F_k)$ will be a specialization of an element of the fraction field of $\Lambda$ of the form

$$\frac{c}{\gamma^r - 1},$$

with $c \in \Lambda$. In any case, we may identify "analytic families of modular forms with values in $\Lambda$" with maps $\mathbf{T} \longrightarrow \Lambda$, provided we allow the zero-th coefficient in the "analytic family" to belong to the fraction field of $\Lambda$. This extends to families with values in $\mathbf{\Lambda}$, for trivial reasons, but it is not clear what happens for more general rings.

We should note the following obvious lemma:

**Lemma III.2.2** *Let* B *be a* $\mathbf{\Lambda}$*-algebra, and assume that the intersection of all the ideals* $P_\chi$B *of* B *is zero. Then a family of parabolic modular forms* $\mathbf{f} : \mathbf{T}_0 \longrightarrow \mathbf{B}$ *with values in* B *is a family of eigenforms if and only if the map* **f** *is a continuous* $\mathbf{Z}_p$-algebra *homomorphism.*

*Proof*: clear. □

Analytic families with values in $\Lambda$ can be detected, as Serre showed in [Se73], by congruence properties. Since $\mathbf{\Lambda}$ is just a sum of copies of $\Lambda$, the same holds for families with values in $\mathbf{\Lambda}$. We refer to Serre's paper for the precise statements: see [Se73, Section 4.4].

Finally, we note that one can use Serre's Eisenstein family to construct any number of examples of $\mathbf{\Lambda}$-valued families of modular forms. For this, let $f$ be a cusp form of $p$-adic weight $(i, k_0)$, and let $\mathrm{E}_k$, with $k \in \mathbf{Z}_p$ denote (the specializations of) Serre's analytic family of Eisenstein series of weight $(0, k)$. Recall that $\mathrm{E}_0 = 1$. Then it is clear (say, from congruence properties) that setting $\mathbf{f}_k = f \cdot \mathrm{E}_{k-k_0}$ defines a family of modular forms with values in $\Lambda$, whose specialization to weight $k_0$ is precisely $f$. We describe this last fact by saying that **f** is a "spreading-out" (over $\Lambda$) of the modular form $f$. Thus, this construction not only provides examples of families of modular forms but also shows that one can spread out any modular form to an analytic family containing it. Unfortunately, the resulting family has few good properties, because the specializations of **f** are not eigenforms for the Hecke algebra, except in very special cases. (For example, in the "rank one" case of Hida's theory of the ordinary part, we can get a family of eigenforms by taking the ordinary part of **f**.) We will later show one way of obtaining a family of eigenforms by twisting a classical modular form by wild characters. Finding other methods for generating families of eigenforms would have significant corollaries for the deformation theory of Galois representations discussed ahead, and is an important open problem.

# III.3 Changing the Level

As we have seen above, the ring **V** of generalized $p$-adic modular functions may be realized as the closure of a union of $\mathbf{Z}_p$-modules of finite rank, by way of divided congruences. In this section, we give another description of **V** in the same spirit, this time taking classical modular forms of constant weight but varying level. The advantage of this description is that it does not necessitate the consideration of congruences of classical forms, so that theorems about classical modular forms over $\mathbf{Z}_p$ can be more easily applied to the situation.

Let us fix a weight $k \geq 2$. We begin by recalling (see Section I.3.5) that we have defined canonical inclusions

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

and by noting that these submodules of $\mathbf{V}$ form a direct system via the canonical inclusions

$$M(B,k,\mathrm{N}p^{\nu}) \hookrightarrow M(B,k,\mathrm{N}p^{\nu+1}).$$

(Note that all of these maps preserve $q$-expansions.) Thus, if we set

$$M(B,k,\mathrm{N}p^{\infty}) = \varinjlim_{\nu} M(B,k,\mathrm{N}p^{\nu}),$$

we get a $q$-expansion preserving inclusion

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

Similarly, by restricting to cusp forms, we may define $S(B,k,\mathrm{N}p^{\infty})$ and get a canonical inclusion preserving $q$-expansions,

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

We are interested in considering the closure of the image (which is, by the $q$-expansion principle, the $p$-adic completion of $S(B,k,\mathrm{N}p^{\infty})$); let us denote it by $\overline{S}(B,k,\mathrm{N}p^{\infty})$. It is clear that the proofs given in Section III.1 allow us to show:

**Proposition III.3.1** *Let $B$ be a $p$-adic ring, and let*

$$\mathrm{h}_k(B,\mathrm{N}p^{\infty}) = \varprojlim_{\nu} \mathrm{h}(B,k,\mathrm{N}p^{\nu}),$$

*where $\mathrm{h}(B,k,\mathrm{N}p^{\nu})$ denotes the Hecke algebra corresponding to the space $S(B,k,\mathrm{N}p^{\nu})$ of cusp forms of weight $k$ and level $\mathrm{N}p^{\nu}$. Let $\mathrm{h}_k(B,\mathrm{N}p^{\infty})$ the inverse limit topology induced by the $p$-adic topology on the $\mathrm{h}(B,k,\mathrm{N}p^{\nu})$. Then the map $f \mapsto \phi_f$ defined as above gives an isomorphism*

$$\overline{S}(B,k,\mathrm{N}p^{\infty}) \xrightarrow{\ \sim\ } \mathrm{Hom}_{\mathbf{Z}_p,conts}(\mathrm{h}_k(\mathbf{Z}_p,\mathrm{N}p^{\infty}),B).$$

*Proof*: Clear; note that the assumption $k \geq 2$ avoids all questions about base change. □

In fact, we have shown nothing new, because of the following result, which was stated without proof in [Hi86a], and whose proof is to appear in a forthcoming paper.

**Theorem III.3.2** [**Hida**] *Let $k \geq 2$ and let $B$ be $p$-adic ring. Then the image of $S(B,k,\mathrm{N}p^{\infty})$ under the canonical inclusion is dense in $\mathbf{V}_{par}(B,\mathrm{N})$. Equivalently, the surjection*

$$\mathbf{T}_0(B,\mathrm{N}) \twoheadrightarrow \mathrm{h}_k(B,\mathrm{N}p^{\infty})$$

*defined by the canonical inclusion is an isomorphism.*

*Proof*: The equivalence of the two statements follows from (III.1.3) and the preceding proposition (and one may even restrict to the case $B=\mathbf{Z}_p$).

*Step 1:* We note, first, that Hida has shown ([Hi86a]) that there are Hecke-invariant inclusions

$$M(B,k,\mathrm{N}p^\infty)\otimes B/p^nB \hookrightarrow M(B,k+1,\mathrm{N}p^\infty)\otimes B/p^nB,$$

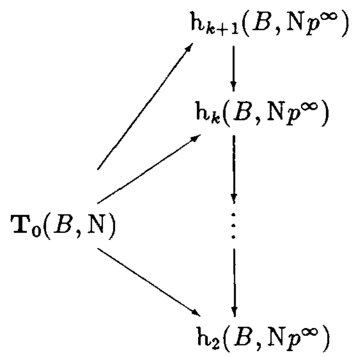
given by multiplication by the appropriate Eisenstein series. These induce Hecke-invariant inclusions

$$\overline{S}(B,k,\mathrm{N}p^\infty) \hookrightarrow \overline{S}(B,k+1,\mathrm{N}p^\infty).$$

Since the Hecke actions extend naturally by continuity to the $p$-adic completions we get surjections

$$\mathfrak{h}_{k+1}(B,\mathrm{N}p^\infty) \twoheadrightarrow \mathfrak{h}_k(B,\mathrm{N}p^\infty).$$

Thus we get a diagram of surjections,




and it follows that it is enough to prove that the map

$$\mathbf{T}_0(B,\mathrm{N}) \twoheadrightarrow \mathfrak{h}_2(B,\mathrm{N}p^\infty)$$

is an isomorphism, or equivalently, that the image of $S(B,2,\mathrm{N}p^\infty)$ is dense in $\mathbf{V}_{par}$.

*Step 2:* The crucial step in the proof is the following result of Hida ([Hi]):

**Theorem III.3.3** [**Hida**] *The maps* $\mathfrak{h}_k(B,\mathrm{N}p^\infty) \longrightarrow \mathfrak{h}_{k-1}(B,\mathrm{N}p^\infty)$ *are all isomorphisms.*

Hida proves this by studying the representation of the Hecke algebras in question on the parabolic cohomology of congruence subgroups of $\mathrm{SL}_2(\mathbf{Z})$, and by invoking a result of Shimura and Ohta. (The result is stated by Shimura in [Shi68] and proved, in the quaternionic case, by Ohta in [Oht82]; using recent results of Harder, one sees that Ohta's proof applies without change in our situation.) We refer the reader to Hida's forthcoming paper.

*Step 3:* The full result now follows easily. Note, first, that Hida's theorem implies that the maps $\overline{S}_k(B,\mathrm{N}p^\infty) \longrightarrow \overline{S}_{k+1}(B,\mathrm{N}p^\infty)$ are isomorphisms. Reducing mod $p^n$, we get that the maps

$$S_k(B/p^nB,\mathrm{N}p^\infty) \longrightarrow S_{k+1}(B/p^nB,\mathrm{N}p^\infty)$$

are isomorphisms, so that we have the following:

**Corollary III.3.4** *Let $n \geq 1$. Fix $k_0 \in \mathbf{Z}$, $k_0 \geq 2$. Then, for any $k \geq 2$ and any classical modular form $f$ of weight $k$ and level $\mathrm{N}p^{\nu}$, there exists $\mu \geq \nu$ and a classical modular form $g$ of weight $k_0$ and level $\mathrm{N}p^{\mu}$ such that $f \equiv g \pmod{p^n}$.*

Now, to conclude the proof, we need to show that the map

$$\mathbf{T}_0 \longrightarrow \mathrm{h}_k(B, \mathrm{N}p^{\infty})$$

is an isomorphism (for any fixed $k$). Since we know it is surjective, we need only show that it is injective. To see this, for each

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

let $P_\chi$ be the ideal of $\mathbf{\Lambda} = \mathbf{Z}_p[[\mathbf{Z}_p^{\times}]]$ generated by

$$\{\langle x, 1\rangle - \chi(x) : x \in \mathbf{Z}_p^{\times}\}.$$

Then it is clear from the duality theorems that $\mathbf{T}_0/P_\chi\mathbf{T}_0$ is dual to the space $\mathsf{S}(B, \chi, \mathrm{N}; 1)$ of $p$-adic cusp forms of weight $\chi$ (because this space is precisely the subspace of $\mathbf{V}_{par}$ consisting of forms satisfying $\langle x, 1\rangle f = \chi(x) f$). We want to look at the map

$$\mathbf{T}_0/P_\chi\mathbf{T}_0 \longrightarrow \mathrm{h}_k(B, \mathrm{N}p^{\infty})/P_\chi \mathrm{h}_k(B, \mathrm{N}p^{\infty}).$$

Dualizing, we get the inclusion

$$\overline{S}_k(B, \mathrm{N}p^{\infty}) \cap \mathsf{S}(B, \chi, \mathrm{N}; 1) \hookrightarrow \mathsf{S}(B, \chi, \mathrm{N}; 1).$$

We claim that this is a surjection. Given $f \in \mathsf{S}(B, \chi, \mathrm{N}; 1)$, we can find a sequence of classical cusp forms $f_n \in S(B, k_n, \mathrm{N})$ such that

$$f(q) \equiv f_n(q) \pmod{p^n},$$

so that $f_n \to f$ in the $q$-expansion topology. By the corollary, we can find $g_n \in S(B, k, \mathrm{N}p^{\nu(n)})$ such that $g_n(q) \equiv f_n(q) \pmod{p^n}$, and then we get that $g_n \to f$, so that $f \in \overline{S}_k(B, \mathrm{N}p^{\infty})$, proving our claim.

Dualizing again, we see that the map

$$\mathbf{T}_0/P_\chi\mathbf{T}_0 \longrightarrow \mathrm{h}_k(B, \mathrm{N}p^{\infty})/P_\chi \mathrm{h}_k(B, \mathrm{N}p^{\infty})$$

is an isomorphism (for any $\chi$). It follows that the map

$$\mathbf{T}_0 \longrightarrow \mathrm{h}_k(B, \mathrm{N}p^{\infty})$$

is injective: if T is in the kernel, then $\mathrm{T} \in P_\chi\mathbf{T}_0$ for all $\chi$; since the divided congruences of $p$-adic cusp forms are dense in $\mathbf{V}_{par}$, we have $\bigcap P_\chi\mathbf{T}_0 = 0$, so that $T = 0$. This proves the theorem. □

## III.4 Deformations of Residual Eigenforms

In [Ma], Mazur has considered the problem of studying the deformations, in a sense to be defined below, of representations of the absolute Galois group of the field of rational numbers in $GL_2(\mathsf{k})$, where $\mathsf{k}$ is a finite field, which he calls a "residual representation". It is well known that some representations of this kind may be obtained from modular forms, and it is an interesting question whether it is true that the deformations of a residual representation arising from a modular form are in some sense modular. In this and the following section, we consider *modular* deformations of a residual representation, i.e., those arising from modular forms, and show that one can construct a universal modular representation which corresponds to a family of $p$-adic modular functions which are eigenforms for the Hecke algebra. We begin, in this section, by considering deformations of "residual eigenforms", i.e., $p$-adic modular functions defined over a finite field $\mathsf{k}$ which are eigenforms for the Hecke algebra. For technical reasons, we restrict the discussion to parabolic modular functions.

**Definition III.4.1** *Let $B$ be a $p$-adic ring. A* Katz eigenform defined over B *is a parabolic generalized modular function which is an eigenform for the Hecke algebra and which is* normalized *in the sense that the coefficient of $q$ in $f(q)$ is equal to 1; we denote the corresponding homomorphism $\mathbf{T}_0 \longrightarrow B$ by $\varphi_f$. This homomorphism induces characters $\chi : \mathbf{Z}_p^\times \longrightarrow B^\times$ and $\varepsilon : (\mathbf{Z}/\mathrm{N}\mathbf{Z})^\times \longrightarrow B^\times$, which we call, respectively, the* weight *and the* nebentypus *of $f$.*

Note that the weight and nebentypus of an eigenform as defined here coincide with the classical weight and nebentypus for classical eigenforms of level N, but *not* for classical eigenforms of level $\mathrm{N}p^\nu$; in the latter case, $\varepsilon$ is the "prime to $p$" part of the nebentypus character. If we need to refer to both terminologies, we will call what we have just defined the "$p$-adic weight" and "$p$-adic nebentypus".

**Definition III.4.2** *A* residual eigenform *is a Katz eigenform $\bar{f}$ defined over a finite field $\mathsf{k}$. Its weight is then necessarily a power of the "Teichmüller" character $\mathbf{Z}_p^\times \longrightarrow \mathbf{F}_p^\times \hookrightarrow \mathsf{k}^\times$ (i.e., the character induced by the canonical map $\mathbf{Z}_p \longrightarrow \mathbf{F}_p$).*

If $\bar{f}$ is a residual eigenform, it follows from Proposition I.3.3 that we have $\bar{f} \in \mathbf{V}_{1,1}$, and hence (because it has a weight) that $\bar{f}$ is the reduction of a classical modular form of level N defined over $W(\mathsf{k})$ (the Witt ring of $\mathsf{k}$) — see [Ka75a]. It then follows from a lemma of Serre and Deligne (see ahead) that $\bar{f}$ is in fact the reduction of a classical *eigenform* of level N defined over some (totally ramified) extension of $W(\mathsf{k})$.

**Definition III.4.3** *A* deformation *of a residual eigenform $\bar{f}$ defined over a finite field $\mathsf{k}$ is a Katz eigenform $f$ defined over an artinian local $W(\mathsf{k})$-algebra $A$ with residue field $\mathsf{k}$ such that $f \mapsto \bar{f}$ under the residue map. More generally, one may take $A$ to be a complete noetherian local $W(\mathsf{k})$-algebra with residue field $\mathsf{k}$ whose quotients by powers of the maximal ideal are artinian.*

Typically, we will be looking at deformations to $W(\mathsf{k})$-algebras $A$ which are either finite extensions of $W(\mathsf{k})$ or their quotients, and at families of such. We will later need to modify this notion somewhat, in order to adjust it to the situation of deformations of Galois representations.

### III.4.1 Universal deformations

Fix a residual eigenform $\bar{f} \in \mathrm{V}(\mathsf{k}, \mathrm{N})$ over a finite field $\mathsf{k}$. Let $\mathcal{C}$ denote the category of artinian local $\mathbf{Z}_p$-algebras $A$ with residue field $\mathsf{k}$, and let $\hat{\mathcal{C}}$ denote the category of complete noetherian local $W(\mathsf{k})$-algebras $B$ with residue field $\mathsf{k}$ such that $B/m^n$ is an object of $\mathcal{C}$ for any power of the maximal ideal $m \subset B$.

Consider the functor $\mathbf{F} : \mathcal{C} \longrightarrow \mathbf{Set}$ defined by

$$\begin{aligned} \mathbf{F}(A) &= \{\text{deformations of } \bar{f} \text{ defined over } A\} \\ &= \{\varphi \in \mathrm{Hom}_{\mathbf{Z}_p-alg}(\mathbf{T}_0, A) \mid \overline{\varphi} = \varphi_{\bar{f}}\} \\ &= \{\varphi \in \mathrm{Hom}_{W(\mathsf{k})-alg}(\mathbf{T}_0(W(\mathsf{k}), \mathrm{N}), A) \mid \overline{\varphi} = \varphi_{\bar{f}}\} \end{aligned}$$

where $\mathrm{Hom}_{W(\mathsf{k})-alg}$ denotes continuous homomorphisms of $W(\mathsf{k})$-algebras (note: $A$ is given the $p$-adic topology), and $\overline{\varphi}$ denotes the composition of $\varphi$ with the residue map $A \longrightarrow \mathsf{k}$. We want to study the representability of this functor.

Since $\bar{f}$ is an eigenform, it corresponds to a continuous algebra homomorphism $\overline{\varphi} : \mathbf{T}_0 \longrightarrow \mathsf{k}$; we denote the kernel of this homomorphism by $\mathfrak{m}$. Let R denote the localization of $\mathbf{T}_0$ at $\mathfrak{m}$, with the induced topology. Then every continuous $\varphi : \mathbf{T}_0 \longrightarrow A$ lifting $\overline{\varphi}$ factors through the canonical map $\mathbf{T}_0 \longrightarrow \mathrm{R}$. This suggests that R will represent our functor; the first difficulty is that R is not complete, so that we need to pass to its completion, which we denote by $\hat{\mathrm{R}}$. Then, of course, for any $p$-adic ring $B$ (complete local noetherian with residue field $\mathsf{k}$), given a continuous algebra homomorphism $\mathbf{T}_0 \longrightarrow B$ lifting $\overline{\varphi}$, we get a continuous algebra homomorphism $\mathrm{R} \longrightarrow B$, which then extends to the completion, giving a continuous algebra homomorphism $\hat{\mathrm{R}} \longrightarrow B$. The converse, however, is not true (unless $\mathsf{k} = \mathbf{F}_p$), because any conjugate of $\bar{f}$ under the Galois group $\mathrm{Gal}(\mathsf{k}/\mathbf{F}_p)$ will determine the same maximal ideal $\mathfrak{m}$, and hence the same completion $\hat{\mathrm{R}}$. In other words, $\hat{\mathrm{R}}$ actually represents the functor

$$\mathbf{G}(A) = \bigcup_{\sigma \in \mathrm{Gal}(\mathsf{k}/\mathbf{F}_p)} \{\text{deformations of } \bar{f}^{\sigma} \text{ defined over } A\}.$$

We prefer to separate the several Galois conjugates; this can be done by base-changing to the Witt ring $W(\mathsf{k})$. Thus, let $\mathfrak{m}$ now denote the kernel of the map $\mathbf{T}_0(W(\mathsf{k}), \mathrm{N}) \longrightarrow \mathsf{k}$ corresponding to $\bar{f}$, and let $\mathrm{R}(\bar{f})$ denote the completion of $\mathbf{T}_0(W(\mathsf{k}), \mathrm{N})$ at $\mathfrak{m}$. Then clearly the deformations of $\bar{f}$ to $B$ correspond precisely to maps $\mathrm{R}(\bar{f}) \longrightarrow B$, and we get:

**Proposition III.4.4** *For any residual parabolic eigenform $\bar{f} \in \mathbf{V}(\mathsf{k},\mathrm{N})$, there exists a complete local $W(\mathsf{k})$-algebra $\mathrm{R}(\bar{f})$ and a family of parabolic modular functions $\mathbf{f}: \mathbf{T}_0(W(\mathsf{k}),\mathrm{N}) \longrightarrow \mathrm{R}(\bar{f})$ which is universal for the deformations of $\bar{f}$; that is, so that for every deformation $f$ defined over a p-adic ring $B$, there exists a continuous homomorphism $\alpha : \mathrm{R}(\bar{f}) \longrightarrow B$ such that $f$ is the specialization of $\mathbf{f}$ via $\alpha$.*

**Remark:** In general, the universal family $\mathbf{f}$ is *not* a modular function defined over $\mathrm{R}(\bar{f})$ (because the map $\mathbf{T}_0(W(\mathsf{k}),\mathrm{N}) \longrightarrow \mathrm{R}(\bar{f})$ is not continuous if we give $\mathrm{R}(\bar{f})$ the $p$-adic topology), so that the functor is not, in the strict sense, representable, in the sense that there is no "universal deformation" (to be precise, the universal deformation is not a modular form, but rather a family of modular forms).

It is interesting to ask if $\mathrm{R}(\bar{f})$ is noetherian. This is equivalent to asking if the "tangent space[2]" of $\mathrm{R}(\bar{f})$ is finite-dimensional, i.e., that

$$\mathbf{F}(\mathsf{k}[\epsilon]) = \mathrm{Hom}_{algebras}(\mathrm{R}(\bar{f}), \mathsf{k}[\epsilon])$$

is a finite-dimensional k-vector space, where $\mathsf{k}[\epsilon] = \mathsf{k} + \mathsf{k}\epsilon$ with $\epsilon^2 = 0$. For example, if $\mathrm{U}\bar{f} = 0$, one can obtain a (not very interesting) deformation to $\mathsf{k}[\epsilon]$ by $g = \bar{f} + \epsilon\,\mathrm{Frob}\,\bar{f}$; this satisfies $\mathrm{U}g = \epsilon g$ and has the same eigenvalues as $\bar{f}$ for the rest of the Hecke algebra. This shows that the dimension of the tangent space is always at least one. Is it finite? We have not been able to answer this in the general situation.

## III.4.2 Deformations outside N$p$

In the next section, we will relate deformations of residual eigenforms and deformations of residual Galois representations. From this point of view, the deformations considered above are not the right thing to consider. The problem is that the representation attached to a modular form depends only on its eigenvalues for the $\mathrm{T}_\ell$ with $\ell \nmid \mathrm{N}p$, i.e., on the eigenvalues for the action of the *restricted* Hecke algebra $\mathbf{T}_0^\star$. Thus, if two Katz eigenforms have the same q-expansion coefficients $a_n$ whenever $n$ is prime to $\mathrm{N}p$, they will determine the same representation. This shows that we should look for deformations up to a weaker notion of equality.

**Definition III.4.5** *Let $B$ be a p-adic ring. We say two Katz eigenforms $f$ and $g$ defined over $B$ are equal outside* N$p$ *if we have $a_n(f) = a_n(g)$ whenever $(n, \mathrm{N}p) = 1$, where $a_n(h)$ denotes the coefficient of $q^n$ in the q-expansion of $h$.*

Since Katz eigenforms are by definition normalized, this is equivalent to requiring that the eigenvalues under $\mathrm{T}_\ell$ be the same for $\ell$ such that $(\ell, \mathrm{N}p) = 1$.

**Definition III.4.6** *Given a residual eigenform $\bar{f}$, we say that a Katz eigenform $g$ defined over $B$ is a deformation of $\bar{f}$ outside* N$p$ *if $a_n(g)$ reduces to $a_n(\bar{f})$ whenever $(n, \mathrm{N}p) = 1$.*

[2] In the sense of Schlessinger in [Sch68]; this is actually the "reduced Zariski tangent space", that is, the Zariski tangent space of $\mathrm{R}(\bar{f})/p\mathrm{R}(\bar{f})$.

We may reinterpret these definitions in terms of our duality theorems: we have $\mathbf{T}_0^\star \subset \mathbf{T}_0$, and we say that two algebra homomorphisms $\mathbf{T}_0 \longrightarrow B$ are equal outside $\mathrm{N}p$ if their restrictions to the subalgebra $\mathbf{T}_0^\star$ coincide. Then, of course, two Katz eigenforms are equal outside $\mathrm{N}p$ if and only if the corresponding algebra homomorphisms are. Similarly, if we denote the homomorphism corresponding to $\bar{f}$ by $\overline{\varphi} : \mathbf{T}_0 \longrightarrow \mathsf{k}$ and its restriction to $\mathbf{T}_0^\star$ by $\overline{\varphi}^\star$, we say a homomorphism $\psi : \mathbf{T}_0 \longrightarrow B$ is a deformation of $\overline{\varphi}$ outside $\mathrm{N}p$ if its restriction to $\mathbf{T}_0^\star$ reduces to $\overline{\varphi}^\star$.

**Definition III.4.7** *We say a parabolic p-adic modular function $f \in \mathbf{V}$ is an eigenform outside* $\mathrm{N}p$ *if it is an eigenform for the action of the restricted Hecke algebra $\mathbf{T}_0^\star$. If $f$ is normalized by $a_1(f) = 1$ and if we denote the continuous map associated to $f$ by $\phi : \mathbf{T}_0 \longrightarrow B$, this is equivalent to requiring that the restriction $\phi^\star$ of $\phi$ to $\mathbf{T}_0^\star$ be an algebra homomorphism*

If we wish to consider deformations of a residual eigenform $\bar{f}$ outside $\mathrm{N}p$, we must of course identify all the maps $\mathbf{T}_0 \longrightarrow B$ which have the same restriction to $\mathbf{T}_0^\star$. This just amounts to considering algebra homomorphisms $\mathbf{T}_0^\star \longrightarrow B$, since it is easy to see that any such may be extended to a $\mathbf{Z}_p$-module homomorphism $\mathbf{T}_0 \longrightarrow B$, by considering the classical case [3].

Let us then consider deformations of $\bar{f}$ as an eigenform outside $\mathrm{N}p$, i.e., modular functions $f$ reducing to $\bar{f}$ outside $\mathrm{N}p$ which are eigenforms under $\mathbf{T}_0^\star$, up to equality outside $\mathrm{N}p$. Equivalently, the question is to describe the continuous algebra homomorphisms $\mathbf{T}_0^\star(W(\mathsf{k}),\mathrm{N}) \longrightarrow B$ lifting the restriction $\overline{\varphi^\star}$ of $\overline{\varphi}$ to $\mathbf{T}_0^\star(W(\mathsf{k}),\mathrm{N})$. Let $\mathbf{m}^\star$ be the kernel of $\overline{\varphi}^\star$. It is clear that this situation is completely analogous to the preceding one, so that the functor $F$ defined by

$$F(B) = \{\text{homomorphisms } \mathbf{T}_0^\star \longrightarrow B \text{ lifting } \overline{\varphi}^\star\}$$

is represented by the completion of $\mathbf{T}_0^\star(W(\mathsf{k}),\mathrm{N})$ at the ideal $\mathbf{m}^\star = \ker(\overline{\varphi}^\star)$. We will denote this ring by $\mathbf{R} = \mathbf{R}(\bar{f})$. Note that the base-change to $W(\mathsf{k})$ is again crucial in order to avoid the problem of conjugation by $\mathrm{Gal}(\mathsf{k}/\mathbf{F}_p)$. Thus, we have:

**Proposition III.4.8** *For any residual parabolic eigenform $\bar{f} \in \mathbf{V}(\mathsf{k},\mathrm{N})$, there exists a complete local $W(\mathsf{k})$-algebra $\mathbf{R}(\bar{f})$ and a map*

$$\mathbf{f} : \mathbf{T}_0^\star(W(\mathsf{k}),\mathrm{N}) \longrightarrow \mathrm{R}(\bar{f})$$

[3]In the classical case, any algebra homomorphism from the restricted Hecke algebra to a $p$-adic ring arises by restriction (not only from a $\mathbf{Z}_p$-module homomorphism, but) from an *algebra* homomorphism from the full Hecke algebra to $B$ (as we remark below, this follows from Atkin-Lehner theory). To put it another way, given a classical modular form which is an eigenform for the action of $\mathbf{T}_0^\star$, one can find an *eigenform* for all of $\mathbf{T}_0$ which "has the same $q$-expansion outside $\mathrm{N}p$". It is interesting to ask if this is still true in the $p$-adic case. The author does not know the answer except for the case $\mathrm{N} = 1$, in which the construction in Section II.3.3 shows that there exists an algebra homomorphism $\mathbf{T}_0 \longrightarrow \mathbf{T}_0^\star$ mapping U to 0, so that the answer is yes.

*which is universal for the deformations outside* $\mathrm{N}p$ *of* $\bar{f}$*; that is, so that for every* $f \in \mathbf{V}(B, \mathrm{N})$ *which is an eigenform outside* $\mathrm{N}p$ *and whose reduction modulo the maximal ideal is equal to* $\bar{f}$ *outside* $\mathrm{N}p$*, there exists a continuous homomorphism* $\alpha : \mathrm{R}(\bar{f}) \longrightarrow B$ *such that the map* $\phi_f^\star : \mathbf{T}_0^\star \longrightarrow B$ *defined by* $f$ *is obtained by* $\phi_f^\star = \alpha \circ \mathbf{f}$.

The ring $\mathbf{R} = \mathbf{R}(\bar{f})$ will have an important role in what follows, especially in relation to constructing modular deformations of a residual representation, and we will need to recall one way to construct it.

For this, recall that $\mathbf{T}_0^\star(W(\mathsf{k}), \mathrm{N})$ is given as an inverse limit of Hecke algebras of finite rank; for any $\nu \geq 1$,

$$\mathbf{T}_0^\star(W(\mathsf{k}), \mathrm{N}) = \varprojlim_j \mathrm{h}_j^\star(W(\mathsf{k}), \mathrm{N}p^\nu),$$

where as above $\mathrm{h}_j^\star(W(\mathsf{k}), \mathrm{N}p^\nu)$ denotes the restricted Hecke algebra corresponding to the space $\mathrm{S}^j(W(\mathsf{k}), \mathrm{N}p^\nu)$ of divided congruences of cuspforms of level $\mathrm{N}p^\nu$ and weight less than or equal to $j$ defined over $W(\mathsf{k})$. Since the topology on $\mathbf{T}_0^\star(W(\mathsf{k}), \mathrm{N})$ is the inverse limit topology, and $\mathsf{k}$ is discrete, the map $\overline{\varphi}^\star$ necessarily factors through some $\mathrm{h}_j^\star(W(\mathsf{k}), \mathrm{N}p^\nu)$, and hence also through any $\mathrm{h}_i^\star(W(\mathsf{k}), \mathrm{N}p^\nu)$ for any $i \geq j$; let $m(j)$ denote the kernel of the homomorphism $\mathrm{h}_j^\star(W(\mathsf{k}), \mathrm{N}p^\nu) \longrightarrow \mathsf{k}$ corresponding to $\overline{\varphi}^\star$, for each $j$ for which it exists. Since these $j$ form an indexing set that is cofinal with the original one, we may as well take the inverse limit only over such $j$, without changing anything. We clearly have $\mathbf{m}^\star = \varprojlim_j m(j)$ and that, after localizing, a $W(\mathsf{k})$-algebra homomorphism

$$R \longrightarrow \mathbf{R} = \varprojlim_j \left(h_j^\star(\mathbf{Z}_p, \mathrm{N}p^\nu)\right)_{m(j)},$$

where $R$ denotes the localization of $\mathbf{T}_0^\star(W(\mathsf{k}), \mathrm{N})$ at the maximal ideal $\mathbf{m}^\star$ and

$$\left(\mathrm{h}_j^\star(\mathbf{Z}_p, \mathrm{N}p^\nu)\right)_{m(j)}$$

denotes the localization of $\mathrm{h}_j^\star(\mathbf{Z}_p, \mathrm{N}p^\nu)$ at the maximal ideal $m(j)$.

**Lemma III.4.9** *The inverse limit*

$$\mathbf{R} = \varprojlim_j \left(h_j^\star(\mathbf{Z}_p, \mathrm{N}p^\nu)\right)_{m(j)}$$

*is the completion of* $R = (\mathbf{T}_0^\star(W(\mathsf{k}), \mathrm{N}))_{\mathbf{m}^\star}$ *with respect to the topology induced by the inverse limit topology on* $\mathbf{T}_0^\star(W(\mathsf{k}), \mathrm{N})$.

*Proof*: This is easy to see, since the kernels of the maps

$$(\mathbf{T}_0^\star(W(\mathsf{k}), \mathrm{N}))_{\mathbf{m}^\star} \longrightarrow \left(\mathrm{h}_j^\star(W(\mathsf{k}), \mathrm{N}p^\nu)\right)_{m(j)}$$

are clearly a basis of neighborhoods of zero in the induced topology. See also [EGA, $0_{\mathrm{I}}$.7.6]. □

To summarize what has been accomplished in this section, given a residual eigenform $\bar{f}$, consider the corresponding algebra homomorphism $\bar{\varphi} : \mathbf{T}_0(W(\mathsf{k}), \mathrm{N}) \longrightarrow \mathsf{k}$ and its restriction $\bar{\varphi}^\star$ to $\mathbf{T}_0^\star(W(\mathsf{k}), \mathrm{N})$; then we may *identify* the three sets

$$\{\text{eigenforms outside } \mathrm{N}p \text{ defined over } B \text{ deforming } \bar{f} \text{ outside } \mathrm{N}p\}/_{\sim},$$

where we set $f \sim g$ if they are equal outside $\mathrm{N}p$,

$$\{W(\mathsf{k})\text{-algebra homomorphisms } \varphi^\star : \mathbf{T}_0^\star(W(\mathsf{k}), \mathrm{N}) \longrightarrow B \text{ lifting } \bar{\varphi}^\star\},$$

and

$$\mathrm{Hom}_{alg}(\mathbf{R}(\bar{f}), B).$$

This allows us to think of the formal spectrum of $\mathbf{R} = \mathbf{R}(\bar{f})$ as the space of eigenforms outside $\mathrm{N}p$ deforming $\bar{f}$. The point of the next section is to identify this last set with the set of modular representations into $\mathrm{GL}_2(B)$ deforming the representation associated to $\bar{f}$. We will do this directly by constructing a representation associated to the "universal" object of the second set given by $\mathbf{T}_0^\star(W(\mathsf{k}), \mathrm{N}) \longrightarrow \mathbf{R}$, which of course corresponds to the identity map of $\mathbf{R}$ in the third set. Before we do so, we need to recall a little of the classical theory of the restricted versus the unrestricted Hecke algebra.

## III.4.3 Some classical results

In the classical situation, the Atkin-Lehner theory of "newforms" tells us that one can find a basis of $S(\overline{\mathbf{Q}}_p, j, \mathrm{N}p)$ (where the bar denotes algebraic closure) composed of "newforms" $f_i$ and of "packages of oldforms"

$$\{V_d(g_i)\},$$

where $g_i$ is a newform of level $\mathrm{N}_1$ dividing $\mathrm{N}p$, $d$ runs over the divisors of $\mathrm{N}p/\mathrm{N}_1$, and $V_d$ denotes the map

$$S(\overline{\mathbf{Q}}_p, j, \mathrm{N}p/d) \longrightarrow S(\overline{\mathbf{Q}}_p, j, \mathrm{N}p)$$

given on $q$-expansions by $q \mapsto q^d$. (If $d \neq p$, $V_d$ gives a map $\mathbf{V}(B, \mathrm{N}/d) \longrightarrow \mathbf{V}(B, \mathrm{N})$; of course, if $d = p$, $V_d$ is just Frob.) The newforms $f_i$ are eigenforms for $\mathbf{T}_0^\star$ "with multiplicity one", and are hence eigenforms for all of $\mathbf{T}_0$. As to the rest, every form in each "package" is an eigenform for $\mathbf{T}_0^\star$, with the *same* eigenvalues (i.e., $\mathbf{T}_0^\star$ acts as scalars). It is *not* always possible to diagonalize the action of $\mathbf{T}_0$ on each package of oldforms (though it is hard to come by an example!), but there is always at least one form in the space generated by the package which *is* an eigenform for $\mathbf{T}_0$ (simply because every matrix has at least one eigenvalue).

Let us call the basis we have been describing the Atkin-Lehner basis of $S(\overline{\mathbf{Q}}_p, k, \mathrm{N}p)$; assume such a basis has been chosen for each weight $k$. Taking a direct sum, we get a basis of

$$\mathsf{S}^k(\overline{\mathbf{Q}}_p, \mathrm{N}p) = \bigoplus_{i \leq k} S(\overline{\mathbf{Q}}_p, i, \mathrm{N}p).$$

Then, if we denote the Hecke algebra of $\mathrm{S}^k(\mathbf{Q}_p, \mathrm{N}p)$ by $\mathfrak{h}_k(\mathbf{Q}_p, \mathrm{N}p)$ and the corresponding restricted Hecke algebra by $\mathfrak{h}_k^\star(\mathbf{Q}_p, \mathrm{N}p)$, the choice of an Atkin-Lehner basis for each weight defines inclusions

$$\mathfrak{h}_k(\mathbf{Q}_p, \mathrm{N}p) \hookrightarrow \bigoplus K_i \oplus \bigoplus M_{n_j}(K_j)$$

and

$$\mathfrak{h}_k^\star(\mathbf{Q}_p, \mathrm{N}p) \hookrightarrow \bigoplus K_i \oplus \bigoplus K_j,$$

where, in each sum, $i$ runs over the newforms, $j$ runs over the packages of oldforms, $n_j$ is the number of forms in the $j^{\text{th}}$ package of oldforms, and $K_i$ and $K_j$ are finite extensions of $\mathbf{Q}_p$. The image of $\mathfrak{h}_k(\mathbf{Q}_p, \mathrm{N}p)$ is a commutative subalgebra of $\bigoplus K_i \oplus \bigoplus M_{n_j}(K_j)$, which may or may not be contained (up to conjugation) in $\bigoplus K_i \oplus \bigoplus (K_j)^{n_j}$, depending on whether the action of $\mathbf{T}_0$ may be diagonalized. It is interesting to consider the images of $\mathfrak{h}_k(\mathbf{Z}_p, \mathrm{N}p)$ and of $\mathfrak{h}_k^\star(\mathbf{Z}_p, \mathrm{N}p)$ under these inclusions. In fact, there is very little that is known beyond the (obvious) fact that these are orders in the images of the Hecke algebras over $\mathbf{Q}_p$, so that, for example, the image of $\mathfrak{h}_k^\star(\mathbf{Z}_p, \mathrm{N}p)$ is contained in a product of discrete valuation rings. This is the fact we will use later. One should note that, though we have stated everything for Hecke algebras over $\mathbf{Z}_p$ and $\mathbf{Q}_p$, the shape of the theory is the same for Hecke algebras over $W(\mathsf{k})$ and its field of fractions (just tensor).

# III.5 Deformations of Galois Representations

We now go on to consider Galois representations and their deformations. As in the preceding chapter, we let $\mathsf{k}$ denote a finite field, and let $\mathcal{C}$ and $\hat{\mathcal{C}}$ be the categories defined above. Let $\mathcal{G}$ denote the absolute Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ of the field of rational numbers.

Let $B$ be a $p$-adic ring. We consider Galois representations "defined over B", i.e., representations

$$\rho : \mathcal{G} \longrightarrow \mathrm{GL}_2(B).$$

We will always assume $\rho$ to be semisimple and unramified outside $\mathrm{N}p$. For any $\ell \nmid \mathrm{N}p$, let $\Phi_\ell$ denote a Frobenius element for $\ell$ in $\mathcal{G}$; since $\rho$ is unramified outside $\mathrm{N}p$, $\mathrm{trace}(\rho(\Phi_\ell))$ and $\det(\rho(\Phi_\ell))$ are well defined. We will say a given Galois representation is *modular* of tame level N if there exists a Katz eigenform $f \in \mathbf{V}(B, \mathrm{N})$ defined over B with $q$-expansion $f(q) = \sum a_i q^i$ (recall that by definition $a_1 = 1$) such that:

$$\mathrm{trace}(\rho(\Phi_\ell)) = a_\ell \quad \text{and} \quad \det(\rho(\Phi_\ell)) = \frac{1}{\ell}\chi(\langle \ell, \ell \rangle),$$

where of course $\chi$ denotes the character of $G(\mathrm{N})$ corresponding to $f$ (its "weight and nebentypus" character). We say that the representation $\rho$ is attached to the eigenform $f$ (though it would be more precise to say it is attached to the system of eigenvalues outside $\mathrm{N}p$ corresponding to $f$). We will say a representation defined over $\mathbf{R}$ is a *family of modular representations* if it is attached, in the same sense as above, to a family of eigenforms $\mathbf{f}$. In this case it is clear that every specialization $\mathbf{R} \longrightarrow A$ will define a modular representation attached to the corresponding specialization of $\mathbf{f}$.

A *residual modular representation* will be a modular representation defined over a finite field $\mathsf{k}$. Since any residual Katz eigenform $f$ must be an eigenform for the diamond operators, we have $f \in \mathbf{V}_{1,1}$, and hence we know that $f$ is a "modular form mod $p$" in the sense of Serre and Swinnerton-Dyer, i.e., that it is the reduction of a classical modular form of level N and some weight $j$. (Since residual eigenforms lift to classical eigenforms (see ahead), we know that a residual modular representation will be unramified outside N$p$.)

It is reasonable to look for modular representations because of the following well-known theorem (whose proof unfortunately is not, as far as this writer is aware, available in complete form anywhere in the literature):

**Theorem III.5.1** *Let $f$ be a classical modular form of level $\mathrm{N}p^{\nu}$ and weight $j \geq 2$ which is an eigenform under the action of the restricted Hecke algebra. Let $\chi$ denote the character describing the action of the diamond operators, and, for each $\ell \not| \mathrm{N}p$, let $a_\ell$ be the eigenvalue corresponding to the Hecke operator $\mathrm{T}_\ell$. Let $K$ be a finite extension of $\mathbf{Q}$ containing the $a_\ell$ and the values of $\chi$. Let $\mathcal{P}$ be a place of $K$, of residual characteristic $p$, and let $K_{\mathcal{P}}$ be the completion of $K$ at $\mathcal{P}$. Then there exists a continuous linear semi-simple representation*

$$\rho_f : \mathcal{G} = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \mathrm{GL}_2(K_{\mathcal{P}}),$$

*which is unramified outside* N$p$ *and satisfies*

$$\mathrm{trace}(\rho_f(\Phi_\ell)) = a_\ell \quad \textit{and} \quad \det(\rho_f(\Phi_\ell)) = \frac{1}{\ell}\chi(\langle \ell, \ell \rangle),$$

*for all $\ell \not| \mathrm{N}p$, where as above $\Phi_\ell$ denotes a Frobenius element for $\ell$.*

*Proof*: This is Theorem 6.1 of [DS74], where one may also find remarks on the extent to which proofs have been published. One should note that, since $\mathcal{G}$ is compact, it is always possible to modify $\rho_f$ so that the image of $\mathcal{G}$ is contained in the ring of integers of $K_{\mathcal{P}}$. □

This shows that, at least for classical eigenforms $f$ defined over valuation rings, one can find a representation $\rho$ attached to $f$. To get the same result for classical *residual* eigenforms $\overline{f}$, one uses a lifting lemma:

**Lemma III.5.2** *Let $\mathcal{H}$ be an arbitrary commutative algebra, $R$ be a discrete valuation ring, and write $R\mathcal{H}$ for $R \otimes_{\mathbf{Z}} \mathcal{H}$. Let $A$ and $B$ be $R\mathcal{H}$-modules, and let $f : A \longrightarrow B$ be a surjective $R\mathcal{H}$-module homomorphism. Let $\Phi : \mathcal{H} \longrightarrow R$ be any map, and assume that there exists $v \in B$ such that $Tv = \Phi(T)v$ for any $T \in \mathcal{H}$. Let $Q$ be a prime ideal in the support of $Rv$. Then there exists a discrete valuation ring $R'$ of finite type over $R$ and a map $\Psi : \mathcal{H} \longrightarrow R'$ such that:*

i. *there exists* $w \in A \otimes_R R'$ *such that* $Tw = \Psi(T)w$ *for any* $T \in \mathcal{H}$, *and*

ii. $\Psi(T) \equiv \Phi(T) \pmod{\mathcal{Q}'}$ *for all* $T \in \mathcal{H}$, *where* $\mathcal{Q}'$ *is the unique prime ideal of* $R'$ *for which* $\mathcal{Q}' \cap R = \mathcal{Q}$.

*Proof*: This is [AS86, Prop. 1.2.2], and it is a generalization of [DS74, Lemma 6.11], which treats the case where $A$ is a free $R$-module of finite rank and $B$ is its reduction modulo the maximal ideal. □

To apply this to our situation, take $A = M(R,k,\mathrm{N})$, $B = M(R,k,\mathrm{N}) \otimes \mathsf{k} = M(R,k,\mathrm{N})$ (provided $k \neq 1$), and let $\mathcal{H}$ be the Hecke algebra corresponding to $A$. We get:

**Corollary III.5.3** *Let* $R$ *be a discrete valuation ring, and let* $\mathfrak{m} \subset R$ *be its maximal ideal. Suppose a classical modular form* $f \in M(R,k,\mathrm{N})$ *is an "normalized eigenform modulo* $\mathfrak{m}$*", i. e., that* $a_1(f) \equiv 1 \pmod{\mathfrak{m}}$ *and that, for any* $\mathrm{T} \in \mathcal{H}$, *we have*

$$\mathrm{T}f \equiv \lambda_{\mathrm{T}} f \pmod{\mathfrak{m}}.$$

*Then there exists a classical modular form* $g \in M(R,k,\mathrm{N})$ *which is an eigenform for the Hecke algebra and which satisfies* $g \equiv f \pmod{\mathfrak{m}}$.

*Proof*: The lemma says that one may find an eigenform $g$ satisfying $\mathrm{T}g = \mu_{\mathrm{T}} g$ with $\mu_{\mathrm{T}} \equiv \lambda_{\mathrm{T}} \pmod{\mathfrak{m}}$. By the classical theory, we may assume $g$ to be normalized, in which case a congruence of eigenvalues implies a congruence of $q$-expansions, and we are done. □

Note that the fact that the modular forms in question are *classical* is crucial to this result, since we need to work with $R$-modules of finite rank.

Now, using the preceding theorem, we get:

**Proposition III.5.4** *Given a residual (normalized) eigenform* $f \in \mathbf{V}(\mathrm{k},\mathrm{N})$, *there exists a residual Galois representation* $\rho$ *unramified outside* $\mathrm{N}p$ *attached to it.*

*Proof*: Since any residual eigenform is necessarily classical of level N, this follows immediately from Lemma III.5.2 and Theorem III.5.1. (See [DS74, Theorem 6.7].) □

We will frequently write $\bar{f}$ for the residual eigenform and $\bar{\rho}$ for the residual representation under consideration, to emphasize that these objects are defined over $\mathsf{k}$. In what follows, $A$ will always denote an object of the category $\hat{\mathcal{C}}$, hence in particular a noetherian local $W(\mathsf{k})$-algebra with residue field $\mathsf{k}$.

We now want to consider deformations of a residual modular representation $\bar{\rho}$ of (tame) level N and defined over k, which we will take as fixed. Since every representation coming into consideration will then be unramified outside $\mathrm{N}p$, let us *redefine* $\mathcal{G}$ to be the Galois group of the maximal extension of $\mathbf{Q}$ unramified outside $\mathrm{N}p$ (i.e., the quotient of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ through which our representations factor).

We follow Mazur's definitions in [Ma], saying that two continuous homomorphisms $\mathcal{G} \longrightarrow \mathrm{GL}_2(A)$ are *strictly equivalent* if they differ by conjugation by an element of $\mathrm{GL}_2(A)$ which is in the kernel of the (residue) map $\mathrm{GL}_2(A) \longrightarrow \mathrm{GL}_2(\mathsf{k})$, and defining a *representation* to be a strict equivalence class of continuous homomorphisms, which we will nevertheless denote by $\rho : \mathcal{G} \longrightarrow \mathrm{GL}_2(A)$.

We will say a Galois representation $\rho : \mathcal{G} \longrightarrow \mathrm{GL}_2(A)$ is a *deformation to A* of a given residual representation $\overline{\rho}$ if any homomorphism $\mathcal{G} \longrightarrow \mathrm{GL}_2(A)$ in the strict equivalence class corresponding to $\rho$ reduces to $\overline{\rho}$ under the canonical map $\mathrm{GL}_2(A) \longrightarrow \mathrm{GL}_2(\mathsf{k})$. (Clearly, this condition depends only on the strict equivalence class, and not on the specific homomorphism.) With these definitions Mazur has shown that *if $\overline{\rho}$ is absolutely irreducible*, there exists a complete noetherian local ring $\mathcal{R}$ with residue field $\mathsf{k}$ (so an object of $\hat{\mathcal{C}}$) and a representation $\rho \longrightarrow \mathrm{GL}_2(\mathcal{R})$ which is a universal deformation of $\overline{\rho}$, i.e., such that *any* deformation of $\overline{\rho}$ defined over a ring A in $\hat{\mathcal{C}}$ is obtained via a (unique) map $\mathcal{R} \longrightarrow A$. (For the construction in much greater generality, and many results about the ring $\mathcal{R}$, see [Ma].)

The point of this section is to construct a universal family of modular deformations of an absolutely irreducible residual representation $\overline{\rho}$, i.e., a representation

$$\rho : \mathcal{G} \longrightarrow \mathrm{GL}_2(\mathrm{R}),$$

where R is some complete noetherian local topological $\mathbf{Z}_p$-algebra in $\hat{\mathcal{C}}$, such that *any* modular deformation of $\overline{\rho}$ defined over a ring $A$ in $\hat{\mathcal{C}}$ is obtained via a (unique) map $\mathrm{R} \longrightarrow A$. We will show that we may take R equal to the ring

$$\mathbf{R} = \mathbf{R}(\overline{f}) = \varprojlim_i \left(\mathrm{h}_i^\star(W(\mathsf{k}), \mathrm{N}p)\right)_{m(i)}$$

considered in the previous section, which is the completion of the localization of the restricted Hecke algebra $\mathbf{T}_0^\star(W(\mathsf{k}), \mathrm{N})$ at the maximal ideal corresponding to the residual eigenform attached to $\overline{\rho}$.

We will construct our representation by using the classical representations, Proposition III.3.2, and a result on representations due to Mazur (in [Ma]). Let $\overline{\rho}$ be a residual modular representation, and let $\overline{f}$ be the associated residual eigenform. *In everything that follows, we assume the residual representation $\overline{\rho}$ to be absolutely irreducible.*

We will now construct a representation of $\mathcal{G}$ defined over the inverse limit $\mathbf{R}$, by constructing a representation defined over each of the localized restricted Hecke algebras at finite levels. As we saw above, the classical theory of newforms implies that we have an inclusion

$$\left(\mathrm{h}_j^\star(W(\mathsf{k}), \mathrm{N}p)\right)_{m(j)} \hookrightarrow \prod \mathcal{O}_i,$$

where the $\mathcal{O}_i$ are complete discrete valuation rings (with residue field k). (Localizing simply chooses some of the valuation rings obtained above, so that the residue fields are all the same and the residue maps all induce $\overline{\varphi}^\star$.) Composing with the projections, we get $W(\mathsf{k})$-algebra homomorphisms

$$\left(\mathrm{h}_j^\star(W(\mathsf{k}), \mathrm{N}p)\right)_{m(j)} \longrightarrow \mathcal{O}_i,$$

which by (III.1.3) and Atkin-Lehner theory (as above) correspond to normalized classical eigenforms of weight less than or equal to $j$ (since a divided congruence which is an eigenform for the diamond operators must be a classical modular form of some weight), and hence to Galois representations $\mathcal{G} \longrightarrow \mathrm{GL}_2(\mathcal{O}_i)$. Taking the product, we get a Galois representation

$$\rho(j) : \mathcal{G} \longrightarrow \mathrm{GL}_2(\prod \mathcal{O}_i),$$

which satisfies, for $\ell \nmid \mathrm{N}p$,

$$\mathrm{trace}(\rho(j)(\Phi_\ell)) = t_\ell \quad \text{and} \quad \det(\rho(j)(\Phi_\ell)) = \frac{1}{\ell}\langle \ell, \ell \rangle,$$

where $t_\ell$ denotes the image of $\mathrm{T}_\ell$ in $(h_j^\star(W(\mathsf{k}), \mathrm{N}p))_{m(j)}$, so that the traces of the representation we have constructed actually take values in the subalgebra $(\mathrm{h}_j^\star(W(\mathsf{k}), \mathrm{N}p))_{m(j)}$ of $\prod \mathcal{O}_i$. It is at this point that we use Mazur's result.

**Theorem III.5.5** *Let $A' \subset A$ be an inclusion of complete semi-local noetherian $W(\mathsf{k})$-algebras, with $A'$ local. Let $A = \prod A_i$ be the factorization of $A$ into a product of local rings, and assume that $A'$ and all the $A_i$ have residue field $\mathsf{k}$. Let $\mathsf{r} : \mathcal{G} \longrightarrow \mathrm{GL}_2(A)$ be a continuous homomorphism such that the induced residual representations $\overline{\mathsf{r}}_i : \mathcal{G} \longrightarrow \mathrm{GL}_2(A_i) \longrightarrow \mathrm{GL}_2(\mathsf{k})$ are all equivalent and absolutely irreducible. Let $\mathrm{H}$ be the image of one of the $\overline{\mathsf{r}}_i$, and suppose that $\mathrm{H}^1(\mathrm{H}, \mathrm{Ad}^0_{\mathrm{H}}) = 0$, where $\mathrm{Ad}^0_{\mathrm{H}}$ denotes the $\mathsf{k}$-vector space $\mathrm{M}_2(\mathsf{k})^0$ of two-by-two matrices of trace zero over $\mathsf{k}$, endowed with the adjoint action of $\mathrm{H}$.*

*Suppose the traces of $\mathsf{r}(g)$ lie in $A'$ for all $g \in \mathcal{G}$. Then there exists a continuous homomorphism $\mathsf{r}' : \mathcal{G} \longrightarrow \mathrm{GL}_2(A')$ such that the representation in $\mathrm{GL}_2(A)$ induced by the inclusion $A' \subset A$ is $A$-equivalent to $\mathsf{r}$.*

*Proof*: This is [Ma, §9 ,Cor. 1′]. □

**Theorem III.5.6** *Suppose that $p \geq 7$. Let $\overline{\rho}$ be an absolutely irreducible modular residual representation defined over $\mathsf{k}$ and of tame level $\mathrm{N}$, and fix notations as above. Let $\mathbf{R}$ denote the inverse limit of the localizations of the restricted Hecke algebras at finite levels at the maximal ideals corresponding to the residual eigenform associated to $\overline{\rho}$:*

$$\mathbf{R} = \varprojlim_n (\mathrm{h}_j^\star(W(\mathsf{k}), \mathrm{N}p))_{m(j)}.$$

*Then there exists a deformation*

$$\rho : \mathcal{G} \longrightarrow \mathrm{GL}_2(\mathbf{R})$$

*of $\overline{\rho}$ to $\mathbf{R}$, satisfying*

$$\mathrm{trace}(\rho(\Phi_\ell)) = \mathrm{T}_\ell \quad \textit{and} \quad \det(\rho(\Phi_\ell)) = \frac{1}{\ell}\langle \ell, \ell \rangle$$

*for any $\ell$ not dividing $\mathrm{N}p$, i.e., which is a family of modular representations of tame level $\mathrm{N}$ attached to the family of eigenforms (outside $\mathrm{N}p$) defined by the canonical homomorphism $\mathbf{T}_0^\star \longrightarrow \mathbf{R}$.*

*Proof*: Given what we have already done, it is only necessary to note that the hypothesis on $\mathrm{H}^1(\mathrm{H}, \mathrm{Ad}^0_{\mathrm{H}})$ is satisfied when $p \neq 5$ (see [CPS]). This gives representations into $\mathrm{GL}_2((h_j(W(\mathsf{k}), \mathrm{N}p))_{m(j)})$ for all sufficiently large $j$. Taking the inverse limit then gives $\rho$; the properties of traces and determinants follow at once from those of the $\rho(j)$, and we are done. □

**Corollary III.5.7** *Under the hypotheses of the Theorem, the ring* $\mathbf{R} = \mathbf{R}(\bar{f})$ *is noetherian.*

*Proof*: It is clearly a quotient of Mazur's ring $\mathcal{R}$ which gives the universal (modular or not) deformation of $\bar{\rho}$, and which is noetherian by construction. □

**Corollary III.5.8** *Let $\bar{f}$ be a residual eigenform whose associated Galois representation is absolutely irreducible, let $A$ be a $p$-adic ring with residue field* $\mathsf{k}$, *and let* $f \in \mathbf{V}_{par}(B, \mathrm{N})$ *be any eigenform outside* $\mathrm{N}p$ *over $A$ which reduces to $\bar{f}$ outside* $\mathrm{N}p$. *Let $\chi$ be the weight-and-nebentypus character of $f$ (i.e., the character giving the action of* $G(\mathrm{N})$*). Then there exists a representation*

$$\rho_f : \mathcal{G} \longrightarrow \mathrm{GL}_2(A)$$

*attached to $f$, i.e., satisfying, for each* $\ell \nmid \mathrm{N}p$,

$$\operatorname{trace}(\rho_f(\Phi_\ell)) = a_\ell \quad \text{and} \quad \det(\rho_f(\Phi_\ell)) = \frac{1}{\ell}\chi(\langle \ell, \ell \rangle).$$

*Proof*: The eigenform $f$ corresponds to a continuous homomorphism $\mathbf{T}_0^\star \longrightarrow A$, which determines a continuous homomorphism $\mathbf{R} \longrightarrow A$ mapping $\mathrm{T}_\ell$ to $a_\ell$ and $\langle \ell, \ell \rangle$ to $\chi(\langle \ell, \ell \rangle)$. Composing this map with $\rho$ gives the desired representation. (Note that this result was previously only known for *classical* eigenforms.) □

Conversely, if a Galois representation deforming an absolutely irreducible residual representation $\bar{\rho}$ to a $p$-adic ring $A$ is attached to an eigenform $f \in \mathbf{V}_{par}(B, \mathrm{N})$, then it is (up to strict equivalence) perforce obtained from $\rho$ via the map $\mathbf{R} \longrightarrow A$ induced by the map $\mathbf{T}_0^\star(W(\mathsf{k}), \mathrm{N}) \longrightarrow A$ corresponding to the eigenform. Thus, $\rho$ is the universal family of (level N) modular deformations of $\bar{\rho}$. We call the ring $\mathbf{R}$ the universal (level N) modular deformation ring of $\bar{\rho}$; when the dependence on $\bar{\rho}$ or on $\bar{f}$ must be made explicit, we will write $\mathbf{R} = \mathbf{R}(\bar{\rho}) = \mathbf{R}(\bar{f})$.

**Caveat:** It is important to note that this construction is strongly dependent on the "tame level". It takes account of all deformations arising from modular forms of level $\mathrm{N}p^\nu$ for any $\nu$, but it does *not* cover those arising, say, from modular forms of level $\mathrm{N}^2p$ (which are still unramified outside $\mathrm{N}p$!). The whole question of the effect of changing the level on our construction is a difficult one, and we will remark further on questions of this type in the conclusion of this chapter.

If we denote Mazur's universal deformation ring by $\mathcal{R} = \mathcal{R}(\overline{\rho})$, it follows that we have a tautological epimorphism

$$\mathcal{R}(\overline{\rho}) \longrightarrow \mathbf{R}(\overline{\rho}).$$

One suspects that this is in fact an isomorphism, i.e., that every deformation of a modular residual representation is modular. We want, in the next section, to obtain information about the modular deformation space and its relation to the total deformation space.

# III.6 The modular deformation space

In [Ma], Mazur has examined in some detail the structure of the space of deformations of a residual representation. His approach is to view a representation defined over a $p$-adic ring $B$ (deforming a given residual representation) as a $B$-valued point in the formal scheme $\mathrm{Spf}(\mathcal{R}(\overline{\rho}))$. The problem then gets translated into studying the geometry of that formal scheme. The surjection $\mathcal{R}(\overline{\rho}) \longrightarrow \mathbf{R}(\overline{\rho})$ should then be viewed as defining the closed formal subscheme $\mathrm{Spf}(\mathbf{R}(\overline{\rho}))$ of modular deformations of the given residual representations, and one would like to obtain more information about this space and its inclusion in the space of all deformations.

## III.6.1 Changing the weight

The first thing to note is that, even staying within the theory of classical modular forms, there are many deformations of any given residual representation. Let $\overline{f}$ be a residual eigenform, and let $\overline{\rho}$ be the attached representation. Assume that $\overline{f}$ is $p$-adically of weight $k$, with $0 \leq k < p-1$, i.e., that it transforms under $(\mathbf{Z}/p\mathbf{Z})^{\times}$ via the $k^{\text{th}}$ power of the Teichmüller character. Then we know (Lemma III.5.2 and the fact that residual eigenforms are necessarily classical) that there exists a classical eigenform $f_i$ of level N and weight $i \equiv k \pmod{p-1}$ lifting $\overline{f}$; choose one such with minimal $i$.

For each $m \geq 0$, we may consider $\overline{f}$ as the reduction of the modular form $f_i \mathrm{E}_{p-1}^m$, which is of weight $j = i + m(p-1)$. Hence, by the lifting lemma (III.5.3), $\overline{f}$ can be lifted to an eigenform of weight $j$, level N, and defined over some extension of $B$, for each such $j$. Thus, we have shown:

**Proposition III.6.1** *Suppose $\overline{f}$ is a residual eigenform which is the reduction of a classical modular $g_i \in M(B, i, \mathrm{N})$ (not necessarily an eigenform). Then, for each $j$ satisfying $j \equiv i \pmod{p-1}$ and $j \geq i$, there exists a finite extension $B_j$ of $B$ and a classical eigenform $f_j \in M(B_j, j, \mathrm{N})$ which reduces to $\overline{f}$ outside* N$p$.

Of course, to each such $f_j$ one can associate a representation $\rho_j$, which will be a deformation of the representation $\overline{\rho}$ attached to $\overline{f}$. This should be viewed as saying that one can vary the deformation "in the direction of the weight".

In fact, one can improve this result a little if one is willing to change the level. Recall, first, that there exists an Eisenstein series $\mathrm{E} \in M(B, 1, \mathrm{N}p)$ (i.e., of weight *one* on $\Gamma_1(\mathrm{N}p)$) satisfying

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

Hence, if $f_i$ is a lifting to weight $i$ and level $\mathrm{N}p$, $\nu \geq 1$ we may multiply by E to get a form of weight $i+1$ and level $\mathrm{N}p$ which still reduces to $\bar{f}$. Applying Corollary III.5.3 once again gives:

**Proposition III.6.2** *Let $\bar{f}$ be a residual eigenform, and assume that there exists a classical modular form $g_i \in M(B, i, \mathrm{N}p)$ reducing to $\bar{f}$. Then, for every $j \geq i$, there exists a finite extension $B_j$ of $B$ and an eigenform $f_j \in M(B_j, j, \mathrm{N}p)$ which reduces to $\bar{f}$.*

The $p$-part of the nebentypus of $f_j$ is determined by the reduction $\bar{f}$, so that we know the precise $p$-adic weight of any deformation to classical forms of level $\mathrm{N}p$: if $\bar{f}$ is of $p$-adic weight $k$, i.e., if we have

$$\langle x, 1\rangle \bar{f} = \omega^k(x)\bar{f}$$

(in $\mathbf{V} \otimes \mathsf{k}$), then $f_j$ must have $p$-adic weight $(k, j)$, i.e.,

$$\langle x, 1\rangle f_j = \omega^k(x)\left(\frac{x}{\omega(x)}\right)^j = \omega^{k-j}(x)x^j.$$

In classical terms, the $p$-part of the nebentypus of a lifting of weight $j$ must be $\omega^{k-j}$.

**Remark:** It is of course possible to state a similar proposition for level $\mathrm{N}p^\nu$, but it is not clear that anything new would be gained. In other words, the forms of which we would be asserting the existence could simply be the $f_j \in M(B', j, \mathrm{N}p)$ (thought of as of level $\mathrm{N}p^\nu$ via the canonical inclusion). To get things which are indeed of higher level, we will need another method, which is the theme of the next subsection.

## III.6.2 Twisting

In the classical theory of modular forms, one encounters the operation of "twisting by a Dirichlet character". If $\chi$ is a character modulo M taking values in a discrete valuation ring $B$ and $f$ is a modular form of level N with $q$-expansion $f(q) = \sum a_n q^n$, this produces a modular form $f^\chi$ of level $\mathrm{NM}^2$ whose $q$-expansion is

$$f^\chi(q) = \sum \chi(n) a_n q^n,$$

where of course we extend $\chi$ to all of $\mathbf{Z}$ in the usual way. (We will recall the definition of $f^\chi$ below.) We will be interested in the case when M is a power of $p$, in which case $\chi$ may simply be thought of as a character of finite order of $\mathbf{Z}_p^\times$, i.e., $\chi : \mathbf{Z}_p^\times \longrightarrow B$ factoring through a quotient $(\mathbf{Z}/p^\nu\mathbf{Z})^\times$ (which we extend to all of $\mathbf{Z}_p$ by setting it equal to zero on $p\mathbf{Z}_p$).

Let $\mathbf{Z}_p^\times = (\mathbf{Z}/p\mathbf{Z})^\times \times \Gamma$ be the usual decomposition of $\mathbf{Z}_p^\times$. We will say that the character $\chi$ is wild if it is trivial on $(\mathbf{Z}/p\mathbf{Z})^\times$, i.e., if it is of order a power of $p$. In that case, its values will we $p$-power roots of unity in $B$, and hence will be congruent to 1 modulo the maximal ideal $\mathfrak{m}$ of $B$, so that we will have $f^\chi(q) \equiv f(q) \pmod{\mathfrak{m}}$. Since twisting by a character transforms eigenforms into eigenforms, this will give us a new source of deformations of a residual eigenform.

The point of this section is to recall the definition in modular terms of the operation of twisting by a character, and show that it can in fact be extended to any character of $\mathbf{Z}_p^\times$, of finite or infinite order. Let us first consider the classical case of a character of finite order. By the $q$-expansion principle, we may extend our ring as necessary, so that we may as well take $B$ to be the completion of the ring of integers in a separable closure of $\mathbf{Q}_p$. A trivial check on $q$-expansions then allows us to determine over which ring our forms turn out to be defined.

Let $\chi$ be a character of $\mathbf{Z}_p^\times$ factoring through $(\mathbf{Z}/p^n\mathbf{Z})^\times$, and let $f \in \mathbf{V}(B, \mathrm{N})$. To define the twist $f^\chi \in \mathbf{V}(B, \mathrm{N})$, we must specify its value on a trivialized elliptic curve with an arithmetic level N structure. Let $\mathrm{E}/B$ be an elliptic curve, $\varphi : \hat{\mathrm{E}} \xrightarrow{\sim} \hat{\mathbf{G}}_m$ be a trivialization, and $\imath : \boldsymbol{\mu}_\mathrm{N} \hookrightarrow \mathrm{E}$ be an arithmetic level N structure. Then $\varphi^{-1}$ determines (by restriction) an inclusion $\boldsymbol{\mu}_{p^n} \hookrightarrow \mathrm{E}$, and we consider the quotient $\mathrm{E}_1 = \mathrm{E}/\varphi^{-1}(\boldsymbol{\mu}_{p^n})$, with its induced trivialization $\varphi_1$ (if $\pi : \mathrm{E} \longrightarrow \mathrm{E}_1$ is the projection, $\varphi_1 = \varphi \circ \check{\pi}$, which is an isomorphism of formal groups because $\check{\pi}$ is étale). The image of $\mathrm{E}[p^n]$ (the kernel of $p^n$ in E) in $\mathrm{E}_1$ is canonically isomorphic to the constant group scheme $\mathbf{Z}/p^n\mathbf{Z}$ (via the Weil pairing), so that we have an isomorphism

$$\mathrm{E}_1[p^n] \cong \boldsymbol{\mu}_{p^n} \times \mathbf{Z}/p^n\mathbf{Z},$$

i.e., what Katz calls an arithmetic $\Gamma(p^n)$ structure on $\mathrm{E}_1$ (see [Ka76, 2.3] for more details).

Suppose now that $\mathrm{H} \subset \mathrm{E}_1$ is an étale subgroup of order $p^n$. Then we may use the isomorphism above to associate to H a $p^n$-th root of unity $\zeta_\mathrm{H}$ (in such a way that we associate 1 to the subgroup $\mathbf{Z}/p^n\mathbf{Z}$). Then we define:

**Definition III.6.3** *Let $f \in \mathbf{V}(B, \mathrm{N})$, and let $\chi$ be a character of $\mathbf{Z}_p^\times$ factoring through $(\mathbf{Z}/p^n\mathbf{Z})^\times$. Then we define the twist of $f$ by $\chi$ by*

$$f^\chi(\mathrm{E}, \varphi, \imath) = \frac{1}{p^n} \sum_{x \bmod p^n} \sum_{\substack{\mathrm{H} \subset \mathrm{E}_1 \\ \#\mathrm{H} = p^n \\ \mathrm{H}\ \text{étale}}} \chi(x) \zeta_\mathrm{H}^{-x} f(\mathrm{E}_1/\mathrm{H}, \varphi_\mathrm{H}, \imath_\mathrm{H}),$$

*where $\mathrm{E}_1$ is as above, and where $\varphi_\mathrm{H}$ and $\imath_\mathrm{H}$ are induced from $\varphi$ and $\imath$ in the obvious way.*

This defines $f^\chi$ as an element of $\mathbf{V}[\frac{1}{p}]$; to show that it is in fact in $\mathbf{V}$, we need only show that its $q$-expansion has integral coefficients. So we must evaluate $f^\chi$ on the Tate curve. Let $\mathrm{E} = \mathrm{Tate}(q)$, and let $\varphi$ and $\imath$ be the canonical trivialization and level N structure. Then we have $\mathrm{E}_1 = \mathrm{Tate}(q^p)$, and the isomorphism obtained above is:

$$\begin{array}{ccc} \boldsymbol{\mu}_{p^n} \times \mathbf{Z}/p^n\mathbf{Z} & \longrightarrow & \mathrm{E}_1[p^n] \\ (\zeta, j) & \mapsto & \zeta q^j, \end{array}$$

where of course we think of $\mathrm{Tate}(q^p)$ as the quotient of $\mathbf{G}_m$ by $q^{p\mathbf{Z}}$. The étale subgroups of order $p^n$ are then $\mathrm{H}_i = <\zeta^i q>$ (the group generated by $\zeta^i q$), where $\zeta$ is a generator of $\boldsymbol{\mu}_{p^n}$, and we simply have $\zeta_{\mathrm{H}_i} = \zeta^i$. Then it is easy to see that $\mathrm{E}_1/\mathrm{H}_i \cong \mathrm{Tate}(q)$, via the map $q \mapsto \zeta^i q$, and this isomorphism is compatible with the canonical trivialization and level N structure. Hence, if $f(q) = \sum a_n q^n$, we have:

$$\begin{aligned} f^{\chi}(\mathrm{E},\varphi,\imath) &= \frac{1}{p^n} \sum_{x \bmod p^n} \sum_{\substack{\mathrm{H}\subset\mathrm{E}_1 \\ \#\mathrm{H}=p^n \\ \mathrm{H}\ \text{étale}}} \chi(x)\zeta_{\mathrm{H}}^{-x} f(\mathrm{E}_1/\mathrm{H},\varphi_{\mathrm{H}},\imath_{\mathrm{H}}) \\ &= \frac{1}{p^n} \sum_{x \bmod p^n} \sum_{i \bmod p^n} \chi(x)\zeta^{-ix} \sum_n \zeta^{in} a_n q^n \\ &= \frac{1}{p^n} \sum_n \sum_{x \bmod p^n} \chi(x) \sum_{i \bmod p^n} \zeta^{i(n-x)} a_n q^n \\ &= \sum_n \chi(n) a_n q^n. \end{aligned}$$

This shows:

**Proposition III.6.4** *Let $f \in \mathbf{V}(B,\mathrm{N})$, and let $\chi : \mathbf{Z}_p^\times \longrightarrow B^\times$ be a character of finite order, which we extend to $\mathbf{Z}_p$ by setting it equal to zero on $p\mathbf{Z}_p$. Assume $f$ has $q$-expansion $f(q) = \sum a_n q^n$; then there exists a $p$-adic modular function $f^\chi \in \mathbf{V}(B,\mathrm{N})$ whose $q$-expansion is given by*

$$f^\chi(q) = \sum \chi(n) a_n q^n.$$

Thus, we have defined, for each character of finite order $\chi$ of $\mathbf{Z}_p^\times$ with values in $B$, a $B$-linear endomorphism

$$\begin{array}{ccc} \mathbf{V}(B,\mathrm{N}) & \longrightarrow & \mathbf{V}(B,\mathrm{N}) \\ f & \mapsto & f^\chi. \end{array}$$

It is immediately clear that we always have $\mathrm{U}f^\chi = 0$ by construction; we will show ahead that whenever $f$ is an eigenform, so is $f^\chi$. First, however, we would like to extend this construction to more general characters, i.e., we would like to show that there exists a twist $f^\chi$ for any character $\mathbf{Z}_p^\times \longrightarrow B$. We do this by noting that the definition of $f \mapsto f^\chi$ can be interpreted as the integral of the character $\chi$ with respect to a certain measure on $\mathbf{Z}_p$ taking values in the space $\mathrm{End}_{\mathbf{Z}_p}(\mathbf{V}(\mathbf{Z}_p,\mathrm{N}))$.

With notations as above, we define a measure $\mu$ on $\mathbf{Z}_p$ as follows: for each $a \in \mathbf{Z}_p$ and each $n \geq 0$, consider the endomorphism $\mu(a,n)$ of $\mathbf{V}$ defined by

$$(\mu(a,n)f)(\mathrm{E},\varphi,\imath) = \frac{1}{p^n} \sum_{\substack{\mathrm{H}\subset\mathrm{E}_1 \\ \#\mathrm{H}=p^n \\ \mathrm{H}\ \text{étale}}} \zeta_{\mathrm{H}}^{-a} f(\mathrm{E}_1/\mathrm{H},\varphi_{\mathrm{H}},\imath_{\mathrm{H}}).$$

Then $\mu(a,n)f$ is clearly an element of $\mathbf{V}[\frac{1}{p}]$; to check that it is in fact in $\mathbf{V}$, we need only compute the effect on $q$-expansions. This is completely analogous to the preceding calculation, and we get: if $f(q) = \sum a_n q^n$, then

$$(\mu(a,n)f)(q) = \sum_{n \equiv a \pmod{p^n}} a_n q^n.$$

This shows:

**Lemma III.6.5** *Let $f \in \mathbf{V} = \mathbf{V}(\mathbf{Z}_p, \mathrm{N})$; then*

*i. for any $a \in \mathbf{Z}_p$ and any $n \geq 0$, $\mu(a,n)f \in \mathbf{V}$, and*

*ii. we have*

$$\mu(a,n)f = \sum_{\substack{b \bmod p^{n+1} \\ b \equiv a \pmod{p^n}}} \mu(b, n+1)f.$$

*Proof*: Given the above computation, the first statement follows immediately from the $q$-expansion principle. It is enough to check the second statement on $q$-expansions, in which case it is obvious. □

Thus, for each $f \in \mathbf{V}$, the assignment

$$a + p^n \mathbf{Z}_p \longmapsto \mu(a,n)f$$

defines a $\mathbf{V}$-valued measure on $\mathbf{Z}_p$; varying $f$ we get that the assignment

$$a + p^n \mathbf{Z}_p \longmapsto \mu(a,n)$$

defines a $\mathrm{End}_{\mathbf{Z}_p}(\mathbf{V})$-valued measure, which we will denote simply by $\mu$. Of course, for any $p$-adic ring $B$, we may think of $\mu$ as taking values in $\mathrm{End}_B(\mathbf{V}(B,\mathrm{N}))$. Then it is clear that, for any character $\chi : \mathbf{Z}_p^\times \longrightarrow B^\times$ of finite order, extended to $\mathbf{Z}_p$ by setting it equal to zero on $p\mathbf{Z}_p$, we have

$$f^\chi = \left( \int_{\mathbf{Z}_p} \chi d\mu \right) f,$$

so that the operation of "twisting by $\chi$" is the integral of $\chi$ with respect to our measure $\mu$. This allows us to extend the idea of twisting by a character to any character $\chi$ of $\mathbf{Z}_p^\times$:

**Definition III.6.6** *Let $\chi : \mathbf{Z}_p^\times \longrightarrow B^\times$ be any homomorphism, and let $f \in \mathbf{V}(B,\mathrm{N})$. Then we define the "twist of $f$ by $\chi$", denoted $f^\chi$, by*

$$f^\chi = \left( \int_{\mathbf{Z}_p} \chi d\mu \right) f,$$

*where as before we extend $\chi$ to $\mathbf{Z}_p$ by $\chi(p\mathbf{Z}_p) = 0$.*

Of course it is just as easy to twist by any function: given a continuous function $\alpha : \mathbf{Z}_p \longrightarrow B$ we write

$$f^{\alpha} = \left(\int_{\mathbf{Z}_p} \alpha d\mu\right) f,$$

and sometimes refer to this as "the twist of $f$ by $\alpha$".

We would now like to understand the effect of twisting by a character on Katz eigenforms (or on eigenforms outside N$p$). As we have already remarked, it is clear (from the $q$-expansions, for example) that we will have $\mathrm{U}(f^{\chi}) = 0$ for *any* $f$. However, this is simply an artifact of our construction (to be precise, it is a consequence of extending $\chi$ by zero to $\mathbf{Z}_p$), and the effect of the Hecke and diamond operators requires analysis. We consider first the case of a character of finite order.

**Lemma III.6.7** *Let $\chi : \mathbf{Z}_p^{\times} \longrightarrow B^{\times}$ be a character of finite order, factoring through $(\mathbf{Z}/p^n\mathbf{Z})^{\times}$. Then we have, for any $f \in \mathbf{V}(B, \mathrm{N})$:*

$$\langle 1, y\rangle(f^{\chi}) = (\langle 1, y\rangle f)^{\chi}, \tag{III.2}$$

*for any $y \in (\mathbf{Z}/\mathrm{N}\mathbf{Z})^{\times}$,*

$$\langle y, 1\rangle(f^{\chi}) = \chi(y)^2(\langle y, 1\rangle f)^{\chi}, \tag{III.3}$$

*for any $y \in \mathbf{Z}_p^{\times}$,*

$$\mathrm{U}(f^{\chi}) = 0, \tag{III.4}$$

*and*

$$\mathrm{T}_{\ell}(f^{\chi}) = \chi(\ell)(\mathrm{T}_{\ell} f)^{\chi}, \tag{III.5}$$

*for any prime $\ell \neq p$.*

*Proof*: Since the definition of twisting by $\chi$ doesn't involve the level N structure $\imath : \boldsymbol{\mu}_{\mathrm{N}} \hookrightarrow \mathrm{E}$ in any way, the first equation is obvious.

For the second equation, recall that the action of the diamond operator $\langle y, 1\rangle$ is given by

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

where we use the canonical action of $\mathbf{Z}_p^{\times}$ on $\hat{\mathbf{G}}_m$. Let, as above, $\mathrm{E}_1 = \mathrm{E}/\varphi^{-1}(\boldsymbol{\mu}_p)$. Then we need to compare the assignment $\mathrm{H} \mapsto \zeta_{\mathrm{H}}$ induced by $\varphi$ to that induced by $y^{-1}\varphi$. The twisting by $y^{-1}$ affects this in two ways: the inclusion $\boldsymbol{\mu}_p \longrightarrow \mathrm{E}_1$ gets twisted by $y^{-1}$, and so does the Weil pairing on $\mathrm{E}[p]$, and hence the identification $\mathrm{E}[p]/\varphi^{-1}(\boldsymbol{\mu}_p) \cong \mathbf{Z}/p\mathbf{Z}$. The net effect is changing $\zeta_{\mathrm{H}}$ to $\zeta_{\mathrm{H}}^{y^{-2}}$. To conclude, we simply calculate, using the multiplicativity of $\chi$:

$$\begin{aligned}
\langle y, 1\rangle(f^{\chi})(\mathrm{E}, \varphi, \imath) &= f^{\chi}(\mathrm{E}, y^{-1}\varphi, \imath) \\
&= \frac{1}{p^n}\sum_{x}\sum_{\mathrm{H}} \chi(x)(\zeta_{\mathrm{H}}^{y^{-2}})^{-x} f(\mathrm{E}_1/\mathrm{H}, y^{-1}\varphi, \imath) \\
&= \chi(y)^2 \frac{1}{p^n}\sum_{x}\sum_{\mathrm{H}} \chi(y^{-2}x)\zeta_{\mathrm{H}}^{-y^{-2}x}(\langle y, 1\rangle f)(\mathrm{E}_1/\mathrm{H}, \varphi, \imath) \\
&= \chi(y)^2(\langle y, 1\rangle f)^{\chi}(\mathrm{E}, \varphi, \imath).
\end{aligned}$$

Hence, we get $\langle y,1\rangle(f^\chi) = \chi(y)^2(\langle y,1\rangle f)^\chi$, as desired. Putting the first two equations together, we get, in particular, that, for any $\ell \not| \mathrm{N}p$,

$$\langle \ell,\ell\rangle(f^\chi) = \chi(\ell)^2(\langle \ell,\ell\rangle f)^\chi.$$

To get the equation for the action of $\mathrm{T}_\ell$, we may either work as above or compute directly on $q$-expansions; we use the latter. Given any modular function $g \in \mathbf{V}(B,\mathrm{N})$, we write $a_n(g)$ for the coefficient of $q^n$ in the $q$-expansion of $g$. Then we know that, if $\ell \not| \mathrm{N}$,

$$a_n(\mathrm{T}_\ell g) = a_{n\ell}(g) + \frac{1}{\ell}a_{n/\ell}(\langle \ell,\ell\rangle g),$$

where we make the convention that $a_{n/\ell}(g) = 0$ if $\ell$ does not divide $n$; if $\ell|\mathrm{N}$, the equation becomes

$$a_n(\mathrm{T}_\ell g) = a_{n\ell}(g).$$

Recall that, for $\chi$ of finite order,

$$a_n(f^\chi) = \chi(n)a_n(f).$$

The case when $\ell|\mathrm{N}$ then follows immediately. For the other case,

$$\begin{aligned}
a_n(\mathrm{T}_\ell f^\chi) &= a_{n\ell}(f^\chi) + \frac{1}{\ell}a_{n/\ell}(\langle \ell,\ell\rangle(f^\chi)) \\
&= a_{n\ell}(f^\chi) + \frac{1}{\ell}a_{n/\ell}(\chi(\ell)^2(\langle \ell,\ell\rangle f)^\chi) \\
&= \chi(n\ell)a_{n\ell}(f) + \frac{1}{\ell}\chi(\ell)^2 a_{n/\ell}((\langle \ell,\ell\rangle f)^\chi) \\
&= \chi(n)\chi(\ell)a_{n\ell}(f) + \frac{1}{\ell}\chi(\ell)^2\chi(n/\ell)a_{n/\ell}(\langle \ell,\ell\rangle f) \\
&= \chi(\ell)\chi(n)\left(a_{n\ell}(f) + \frac{1}{\ell}a_{n/\ell}(\langle \ell,\ell\rangle f)\right) \\
&= \chi(\ell)a_n((\mathrm{T}_\ell f)^\chi).
\end{aligned}$$

By the $q$-expansion principle, it follows that $\mathrm{T}_\ell(f^\chi) = \chi(\ell)(\mathrm{T}_\ell f)^\chi$, and we are done. □

If we examine the calculations above, it is clear that we have only used the fact that $\chi : \mathbf{Z}_p \longrightarrow B$ is a (locally constant) multiplicative function. Let $\alpha : \mathbf{Z}_p \longrightarrow B$ be any multiplicative function, so that $\alpha(xy) = \alpha(x)\alpha(y)$ for any $x, y \in \mathbf{Z}_p$. If we approximate $\alpha$ by locally constant functions $\alpha_n$, these will satisfy

$$\alpha_n(xy) \equiv \alpha_n(x)\alpha_n(y) \pmod{p^{\nu(n)}}$$

for some $\nu(n)$, so that the calculations above all go through after changing some of the equalities to congruences. Taking the limit, we get:

**Corollary III.6.8** *Let $\alpha : \mathbf{Z}_p \longrightarrow B$ be any continuous multiplicative function, and let $f \in \mathbf{V}(B,\mathrm{N})$ with $q$-expansion $f(q) = \sum a_n q^n$. Then:*

*i.* $(f^\alpha)(q) = \sum \alpha(n) a_n q^n$,

*ii.* $\langle 1, y\rangle(f^\alpha) = (\langle 1, y\rangle f)^\alpha$, *for any* $y \in (\mathbf{Z}/\mathrm{N}\mathbf{Z})^\times$,

*iii.* $\langle y, 1\rangle(f^\alpha) = \alpha(\ell)^2(\langle y, 1\rangle f)^\alpha$, *for any* $y \in \mathbf{Z}_p^\times$,

*iv.* $\mathrm{U}(f^\chi) = \alpha(p)(\mathrm{U}f)^\alpha$ *and*

*v.* $\mathrm{T}_\ell(f^\alpha) = \alpha(\ell)(\mathrm{T}_\ell f)^\alpha$, *for any prime* $\ell \neq p$.

In particular, we get:

**Corollary III.6.9** *Let* $\chi : \mathbf{Z}_p^\times \longrightarrow B^\times$ *be any character, extended to* $\mathbf{Z}_p$ *by* $\chi(p\mathbf{Z}_p) = 0$, *and let* $f \in \mathbf{V}(B, \mathrm{N})$ *be a Katz eigenform with weight-and-nebentypus character* $\varepsilon : G(\mathrm{N}) \longrightarrow B^\times$ *and with* $\mathrm{T}_\ell f = \lambda_\ell f$ *for any prime* $\ell \neq p$. *Then* $f^\chi$ *is a Katz eigenform with weight-and-nebentypus character* $\varepsilon\chi^2$, *satisfying* $\mathrm{U}f^\chi = 0$ *and* $\mathrm{T}_\ell(f^\chi) = \chi(\ell)\lambda_\ell f^\chi$.

For some special functions $\alpha : \mathbf{Z}_p \longrightarrow B$, the operation of twisting by $\alpha$ turns out to be a well-known operator on modular functions. For a first example, let $\chi : \mathbf{Z}_p^\times \longrightarrow \mathbf{Z}_p^\times$ be the trivial character: $\chi(x) = 1$ for all $x \in \mathbf{Z}_p^\times$. Let $f \in \mathbf{V}$ have $q$-expansion $f(q) = \sum a_n q^n$. Then

$$f^\chi(q) = \sum_{(n,p)=1} a_n q^n = (f - \mathrm{Frob}(\mathrm{U}f))(q),$$

so that twisting by the trivial character is the same as $1 - \mathrm{Frob} \circ \mathrm{U}$. For a more interesting example, let $\alpha : \mathbf{Z}_p \longrightarrow \mathbf{Z}_p$ be the identity function. Then we get

$$f^\alpha(q) = \sum n a_n q^n = q \frac{d}{dq} f(q),$$

so that twisting by the identity is the same as the "$q\frac{d}{dq}$" operator considered by Serre (in [Se73]) and Katz (in [Ka76]). Since $\alpha$ is multiplicative, it follows from the lemma that $q\frac{d}{dq}$ "is of $p$-adic weight 2", i.e., sends modular forms of weight $(i, k)$ to forms of weight $(i + 2, k + 2)$, which is a result obtained by Katz in [Ka76].

### III.6.3 Families of twists, and an estimate for the Krull dimension of the modular deformation ring

We can now use the twisting process defined above to produce one-parameter families of deformations of a residual eigenform (or, equivalently, of a modular residual deformation). If we begin with a Katz eigenform $f$ with weight-and-nebentypus character $\varepsilon : \mathbf{Z}_p^\times \longrightarrow B^\times$, then twisting by a character $\chi : \mathbf{Z}_p^\times \longrightarrow B^\times$ produces another Katz eigenform $f^\chi$, with weight-and-nebentypus character $\epsilon\chi^2$. In particular, if $\chi$ is wild, i.e., if $\chi(x)$ is a one-unit in $B^\times$ for all $x \in \mathbf{Z}_p^\times$, then we will have $f^\chi \equiv f \pmod{\mathfrak{m}}$, where $\mathfrak{m}$ is the maximal ideal of $B$. Thus, if $f$ is a deformation of a residual eigenform $\bar{f}$, we have constructed a family of deformations of $\bar{f}$ indexed by the wild characters $\mathbf{Z}_p^\times \longrightarrow B^\times$, where of course we may vary $B$. In what follows, we will always assume $B$ to be a finite extension of $\mathbf{Z}_p$.

To actually construct a family of functions (as defined above) recall the identifications of $\mathbf{V}(B,\mathrm{N})$ with $\mathrm{Hom}_{\mathbf{Z}_p,cont}(\mathbf{T}_0(W(\mathsf{k}),\mathrm{N}),B)$ (Section III.1) and of $\Lambda_B = B[[\Gamma]]$ with the space of $B$-valued measures on $\Gamma$ (see, e.g., [Se73]). Let $\mathbf{\Lambda}_B = B[[\mathbf{Z}_p^\times]] = \Lambda_B[(\mathbf{Z}/p\mathbf{Z})^\times]$, which we of course can identify with the space of $B$-valued measures on $\mathbf{Z}_p^\times$. Given a Katz eigenform $f \in \mathbf{V}(B,\mathrm{N})$, consider the $\mathbf{V}(B,\mathrm{N})$-valued measure on $\mathbf{Z}_p^\times$ given by (the restriction of) $\mu(f)$. For each $\mathrm{T} \in \mathbf{T}_0$, evaluation at T gives a $B$-valued measure on $\mathbf{Z}_p^\times$, hence an element of $\mathbf{\Lambda}_B$. This defines a continuous homomorphism $\tau_f : \mathbf{T}_0(W(\mathsf{k}),\mathrm{N}) \longrightarrow \mathbf{\Lambda}_B$, and hence a family of modular forms. In explicit terms, $\tau_f$ can be described as follows: if

$$\phi_f : \mathbf{T}_0 \longrightarrow B$$

is the canonical homomorphism corresponding to $f$, we have:

i. $\tau_f(\langle 1,y\rangle) = \phi_f(\langle 1,y\rangle)$,

ii. $\tau_f(\langle y,1\rangle) = \phi_f(\langle y,1\rangle)\langle y,1\rangle^2$,

iii. $\tau_f(\mathrm{U}) = 0$, and

iv. $\tau_f(\mathrm{T}_\ell) = \phi_f(\mathrm{T}_\ell)\langle \ell,1\rangle$.

Then, for any character $\chi$, $f^\chi$ is just the specialization of the family $\tau_f$ via $\chi$ (and, in fact, the result that $\tau_f$ is a continuous homomorphism follows at once from this fact, which is obvious from the $q$-expansions). If $f$ is a deformation of a residual eigenform $\bar{f}$, and if we restrict to wild characters, which of course are trivial on $(\mathbf{Z}/p\mathbf{Z})^\times \subset \mathbf{Z}_p^\times$, we get a continuous homomorphism $\mathbf{T}_0(W(\mathsf{k}),\mathrm{N}) \longrightarrow \Lambda_B$, i.e., a family of modular forms, all of which reduce to $\bar{f}$ outside $\mathrm{N}p$ (because $\chi(x) \equiv 1 \pmod{p}$ for all $x$. Thus, we get a continuous homomorphism $\tau_f^\star : \mathbf{T}_0^\star(W(\mathsf{k}),\mathrm{N}) \longrightarrow \Lambda_B$ which defines a continuous homomorphism $\mathbf{R} \longrightarrow \Lambda_B$ (because every $f^\chi$ reduces to $\bar{f}$). Since $\Lambda_B \cong B[[T]]$, this should be thought of as a one-dimensional analytic family of deformations of $\bar{f}$.

To summarize:

**Proposition III.6.10** *Given any residual eigenform $\bar{f}$ of level* N*, and given any Katz eigenform $f \in \mathbf{V}(B,\mathrm{N})$ reducing to $\bar{f}$ modulo the maximal ideal of $B$, there exists a one-dimensional analytic family of deformations*

$$\tau_f^\star : \mathbf{R} \longrightarrow \Lambda_B$$

*giving the twists of $f$ by wild characters of $\mathbf{Z}_p^\times$.*

In terms of representations, if $\bar{\rho}$ is the residual representation attached to $\bar{f}$, we get, by composing the universal modular deformation $\rho : \mathcal{G} \longrightarrow \mathrm{GL}_2(\mathbf{R})$ we get, for every modular deformation $\rho_f : \mathcal{G} \longrightarrow \mathrm{GL}_2(B)$, a one-dimensional analytic family of deformations $\mathcal{G} \longrightarrow \mathrm{GL}_2(\Lambda_B)$, whose specialization under any character $\chi : \Gamma \longrightarrow B^\times$ is simply the twist of $\rho_f$ by the one dimensional representation of $\mathcal{G}$ given by $\chi$ in the obvious way. (See Mazur's paper [Ma] for more on twisting representations.) Thus we have shown that any twist of a modular deformation of $\bar{f}$ by a wild character (i.e., a character of $\Gamma$) is again a modular deformation of $\bar{f}$. We will use this to show that the Krull dimension of the universal modular deformation ring is at least 3. For this, we need two easily-proved facts:

**Lemma III.6.11** *Let $f \in \mathbf{V}(B, \mathrm{N})$ be a deformation of a residual eigenform $\bar{f}$, and let $\tau_f^\star : \mathbf{R}_B = \mathbf{R}\hat{\otimes}B \longrightarrow \Lambda_B$ be the map induced by the family of twists of $f$ defined above. Then $\tau_f^\star$ is a continuous surjective homomorphism of $\mathbf{Z}_p$-algebras.*

*Proof*: Everything but surjectivity has already been noted. To get surjectivity, it is enough to check that, for any $\gamma \in \Gamma$, the element $\langle\gamma\rangle$ is in the image (where, as above, we use angular brackets to distinguish elements of $\Gamma$ from themselves as elements of $\mathbf{Z}_p^\times$). But, since 2 is invertible in $\mathbf{Z}_p$ and so is the image $\gamma$ under the character corresponding to $f$, this is immediate from the formula $\tau_f^\star(\langle y, 1\rangle) = \phi_f(\langle y, 1\rangle)\langle y, 1\rangle^2$ above. In fact, this shows that the composite map $\Lambda_B \longrightarrow \mathbf{R}_B \longrightarrow \Lambda_B$ is already surjective. □

**Lemma III.6.12** *Let $f \in \mathbf{V}(B, \mathrm{N})$ be a deformation of a residual eigenform $\bar{f}$, and assume $f$ is classical, i.e., $f \in M(B, k, \epsilon, \mathrm{N}p^\nu)$ for some $\nu$ and appropriate nebentypus $\epsilon$. Then*

*i. the twist of $f$ by any character of finite order is again classical; specifically, if $\chi : \mathbf{Z}_p^\times \longrightarrow B^\times$ factors through $(\mathbf{Z}/p^n\mathbf{Z})^\times$, we have $f^\chi \in M(B, k, \epsilon\chi^2, \mathrm{N}p^{\nu+2n})$*

*ii. the twist of $f$ by any character $\chi : \mathbf{Z}_p^\times \longrightarrow B^\times$ of infinite order is* not *a classical modular form.*

*Proof*: The first claim, is, of course, well-known, and obvious from the definitions. For the second claim, we must use considerations of $p$-adic Hodge structure of the corresponding representations. One knows that, in the classical case, the $p$-adic Hodge twists (in the sense of [MW86]) of a representation attached to a classical form of weight $k$ whose reduction is absolutely irreducible are $(0, k-1)$ (see [MW86]); twisting by the character $\gamma \mapsto \gamma^j$ gives a representation with twists $(j, j+k-1)$, which is therefore not classical. □

Then we get:

**Proposition III.6.13** *The Krull dimension of the universal modular deformation space $\mathbf{R} = \mathbf{R}(\bar{f}) = \mathbf{R}(\bar{\rho})$ is at least three.*

*Proof*: Let $k$ be the least weight for which there is a classical lift $f_k$ of level N. Then for each $j \geq k$, $j \equiv k \pmod{p-1}$, there exists a classical lift $f_j$ of weight $j$ and level N. Then, for each such $j$, we have obtained a continuous surjective homomorphism

$\tau_f^\star : \mathbf{R} \longrightarrow \Lambda_{B_j}$ (where we $B_j$ is a finite extension of $\mathbf{Z}_p$). Since the dimension of $\Lambda_{B_j}$ is two, we need only prove that the kernel of one (and hence of any) of these maps is not a minimal prime ideal of $\mathbf{R}$. However, the preceding corollary shows that these ideals are all distinct; since $\mathbf{R}$ is noetherian, they cannot be minimal primes (because there is only a finite number of such), and we are done. □

It would be interesting to get a more precise estimate of the Krull dimension of the modular deformation ring. For example, Mazur and Boston have considered the case of a residual representation

$$\overline{\rho} : \mathcal{G} \longrightarrow S_3 \subset \mathrm{GL}_2(\mathbf{F}_p),$$

where $S_3$ is the symmetric group on three letters. Under some additional hypotheses, they show that the universal deformation space $\mathcal{R}(\overline{\rho})$ has Krull dimension four (in fact, that it is a power-series ring in three variables over $\mathbf{Z}_p$). Is the modular deformation ring also of dimension four in this case?

## III.6.4 The ordinary case

In the case where the residual representation is attached to an ordinary modular form, one can consider the "universal ordinary deformation", i.e., the universal deformation associated to ordinary modular forms. Using the work of Hida in [Hi86a], Mazur and Wiles (in [MW86]) constructed the universal ordinary deformation (under some restrictive hypotheses) and determined several of its properties, especially with respect to the image of the decomposition group at $p$. In this section, we summarize these results and point out their relation to our larger deformation space.

For simplicity, and to agree with the situation in [MW86], let us assume that the level $\mathrm{N} = 1$. Let $\overline{f} \in \mathbf{V}_{par}(\mathsf{k}, 1)$ be an ordinary residual parabolic eigenform, and let $\overline{\rho}$ be the attached residual representation. By duality, $\overline{f}$ corresponds to a map $\mathbf{T}_0 \longrightarrow \mathsf{k}$, which necessarily factors through the ordinary Hecke algebra $\mathbf{T}_0^{ord} = e_0\mathbf{T}_0$. In fact, we can say more; a residual eigenform has a weight $\chi : \mathbf{Z}_p^\times \longrightarrow \mathsf{k}$, which, as we remarked above, must be a power of the Teichmüller character: $\chi = \omega^i$. Then it is clear that the map $\phi_{\overline{f}} : \mathbf{T}_0 \longrightarrow \mathsf{k}$ corresponding to $\overline{f}$ must in fact factor through the summand $\mathsf{T}_i$ of $\mathbf{T}_0^{ord}$ defined in the Appendix to the last chapter. The various Hecke algebras and the map defined by $\overline{f}$ fit together like this:

$$\mathbf{T}_0^\star \subset \mathbf{T}_0 \xrightarrow{e_0} \mathbf{T}_0^{ord} \longrightarrow \mathsf{T}_i \longrightarrow \mathsf{k}.$$

Let $\mathbf{m}^\star$, $\mathbf{m}$, $\mathbf{m}^{ord}$, and $\mathfrak{m}$ denote the kernels of $\phi_{\overline{f}}$ in each of the Hecke algebras.

The construction of Mazur and Wiles requires a hypothesis on the weight $i$. Hence *assume, for the remainder of this section, that* $i \neq 2$. In terms of the representation $\overline{\rho}$, this means that $det(\overline{\rho})$ is not the cyclotomic character.

In this situation, Mazur and Wiles construct a representation

$$\rho^{ord} : \mathcal{G} \longrightarrow \mathrm{GL}_2(\mathsf{R}),$$

where $\mathsf{R}$ denotes the completion of the Hecke algebra $\mathsf{T}_i$ at the maximal ideal $\mathsf{m}$, which gives the universal ordinary deformation of the representation attached to $\bar{f}$, in the same sense as before: any representation deforming $\bar{\rho}$ which is attached to an ordinary $p$-adic modular form is obtained from $\rho^{ord}$ via the induced map from $\mathsf{R}$ to the ring of definition of the attached modular form. They then obtain theorems about the action the decomposition and the inertia groups at $p$, including a description of the $p$-adic Hodge twists of the specializations. We refer to [MW86] for further details. One should note that Mazur and Wiles have also obtained a necessary condition, in terms of the action of the inertia group at $p$, representation to be attached to an ordinary modular form. This condition is conjectured to be sufficient, in which case it would give representation-theoretic description of the ordinary deformation space. See the discussion in [MW86] and [Ma] for more details.

We would like to compare the universal ordinary deformation to the universal modular deformation we constructed before. It is clear, from the universal property of the constructions, that there must exist a map $\mathbf{R} \longrightarrow \mathsf{R}$ so that $\rho^{ord}$ is the representation obtained from $\boldsymbol{\rho}$ via this map; it is also clear (compare the traces of the Frobenii) that this map must be the map induced by the canonical maps

$$\mathbf{T}_0^\star \hookrightarrow \mathbf{T}_0 \longrightarrow \mathsf{T}_i$$

by completion at the maximal ideal. From the analysis in [MW86], we can show:

**Proposition III.6.14** *Let $\bar{f}$ be an ordinary residual eigenform of p-adic weight $i \neq 1$ and level $\mathrm{N} = 1$. Let $\mathbf{R} = \mathbf{R}(\bar{f})$ be the universal modular deformation ring and let $\mathsf{R} = \mathsf{R}(\bar{f}) = (\mathsf{T}_i)_{\mathsf{m}}$ be the completion of the $i^{\mathrm{th}}$ component of the ordinary Hecke algebra at the ideal corresponding to $\bar{f}$. Consider the map $\mathbf{R} \longrightarrow \mathsf{R}$ induced from the composition of the inclusion with the projection on $\mathsf{T}_i$. Then $\mathsf{R}$ is at most a quadratic extension of the image of* $\mathbf{R}$.

*Proof*: Since we are assuming $\mathrm{N} = 1$, it is clear that $\mathsf{R}$ is generated over the image of $\mathbf{R}$ by (the image in the completion of) the U operator. To see that U is quadratic over the image of $\mathbf{R}$, we show that for some unit $\lambda \in \Lambda$, the element $\mathrm{U} + \lambda \mathrm{U}^{-1} \in \mathsf{R}$ belongs to the image of $\mathbf{R}$. (Note that U is invertible in $\mathsf{R}$, by definition of the ordinary part.) This, however, follows at once from the results in [MW86, §8] mentioned above: let $\sigma$ be an element of the decomposition group at $p$ mapping modulo the inertia group to a generator of $\hat{\mathbf{Z}} = \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$; then we have

$$\mathrm{trace}(\rho^{ord}(\sigma)) = \mathrm{U} + \lambda \mathrm{U}^{-1},$$

where $\lambda$ is some unit of $\Lambda$; therefore, for any such $\sigma$,

$$\mathrm{trace}(\boldsymbol{\rho}(\sigma)) \mapsto \mathrm{U} + \lambda \mathrm{U}^{-1} \in \mathsf{R},$$

so that $\mathrm{U} + \lambda \mathrm{U}^{-1}$ is in the image of $\mathbf{R}$, as desired. □

Since R is a flat $\Lambda$-algebra of finite rank (see Section II.4), its Krull dimension is two. Taking all possible twists gives a map

$$\tau^{ord} : \mathbf{R} \longrightarrow \mathsf{R}[[\Lambda]],$$

as in the preceding section (because a twist of an ordinary modular form is never ordinary). This last ring has Krull dimension three, so that we need to study the kernel $\mathcal{I} = \ker(\tau^{ord})$; if this is not contained in a minimal prime ideal of $\mathbf{R}$, it will follow that the dimension of $\mathbf{R}$ is at least four. In the example worked out by Mazur and Boston ("neat $S_3$ extensions"), this would suffice to show that every deformation is modular.

We have not been able to prove that the dimension of the modular deformation space is at least four. In any case, it is easy to see that the modular deformation space is always strictly larger than the ordinary deformation space, by showing that there exist deformations of the representation $\overline{\rho}$ attached $\overline{f}$ which are modular, but whose corresponding modular forms are neither ordinary nor twists of ordinary modular forms. Assume $k \geq 3$, $k \equiv i \pmod{p-1}$. Then $\overline{f}$ can be lifted to a classical modular form of weight $k$ which can be assumed to be an eigenform and is necessarily ordinary (because it is congruent to $\overline{f}$). For deforming the *representation*, however, all we need is to find a lift that is congruent to $\overline{f}$ *outside* $\mathrm{N}p$; we will show that if the weight is high enough, one can always find a deformation outside $\mathrm{N}p$ which is not ordinary, but is classical of level N, so that it is certainly not a twist of a classical modular form.

Let $f$ be any lift of $\overline{f}$ to an (ordinary) eigenform of weight $k$ and level N. Assume for simplicity that $\mathsf{k} = \mathbf{F}_p$ and therefore that $f$ is defined over some totally ramified extension $B$ of $\mathbf{Z}_p$. Consider the modular form

$$g_1 = \mathrm{E}_{p-1}^{j}(\mathrm{E}_{p-1}^{k} f - (\mathrm{T}_p f)^p) \in M(B, pk + j(p-1), \mathrm{N}).$$

It is clear that the reduction of $g_1$ is an eigenform (since it is congruent, modulo the maximal ideal of $B$, to $f - \mathrm{Frob}(\mathrm{U}f)$); thus, by Lemma III.5.2, there must exist an eigenform $g \in M(B, pk + j(p-1), \mathrm{N})$ (possibly after base-change) which is congruent to $g_1$. Then $g$ will be a deformation of $\overline{f}$ outside $\mathrm{N}p$, it will be classical of weight $pk + j(p-1)$ and level N, and it will not be ordinary. Hence the representation attached to $g$ will not be obtained by specializing the universal ordinary representation $\rho^{ord}$.

Thus, we have shown:

**Proposition III.6.15** *Let $\overline{f}$ be an ordinary residual eigenform defined over* k. *Suppose there exists a classical eigenform $f$ of weight $k_0$ and level* N *defined over a $W(\mathsf{k})$-algebra $B$ reducing to $\overline{f}$ modulo the maximal ideal, and hence ordinary:* $\mathrm{T}_p f \not\equiv 0 \pmod{\mathfrak{m}}$. *Then, for any weight $k \geq pk_0$, $k \equiv k_0 \pmod{p-1}$, there exists a classical eigenform $g$ of weight $k$ and level* N *defined over a finite extension $B'$ of $B$ (which will depend on $k$) whose reduction modulo the maximal ideal is equal to $\overline{f}$ outside* $\mathrm{N}p$ *but which is* not *ordinary:* $\mathrm{T}_p g \equiv 0 \pmod{\mathfrak{m}}$.

**Remark:** To be absolutely precise, $f$ itself is not ordinary, because it is not an eigenform under the U operator. However, its ordinary projection $e_0 f$ is nonzero and equal to $f$ outside $p$, while $e_0 g = 0$. The point is simply that for each large enough weight, there is always at least one ordinary and one non-ordinary deformation which is not the twist of any ordinary deformation.

**Corollary III.6.16** *If the residual representation $\overline{\rho}$ is attached to an ordinary residual eigenform $\bar{f}$, the modular deformation space of $\overline{\rho}$ is strictly larger than the ordinary deformation space, even if we add all the twists of ordinary deformations.*

Since the Krull dimension of the ordinary-plus-twists deformation ring is three, one is led to ask:

**Question III.1** *Suppose that the residual modular representation $\overline{\rho}$ is attached to an* ordinary *modular form. Is the Krull dimension of the universal modular deformation ring always greater than or equal to four?*

At least in the case when $\bar{f}$ is ordinary, the preceding discussion suggests that the answer may be yes; in the "neat $S_3$" case considered by Mazur and Boston, this would imply that every deformation of the residual representation in question is attached to a $p$-adic modular form.

## III.7 Further Questions

This final section collects some of the questions which seem to arise in relation to the topics we have discussed in this book, and that may suggest paths for further research. When appropriate, we have suggested what we suspect will be the answer, but we have not dignified these suspicions by calling them conjectures, since for the most part there is little evidence one way or the other. We use the notations and conventions defined above, giving references only when necessary.

To begin with, there are several questions associated to the subject of Chapter II: the U operator and its eigenforms. As we saw, there exists, for each $\alpha \geq 0$ a "slope $\alpha$ projection"

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

defining a splitting

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

where $M^{(\alpha)}$ is the "slope $\alpha$ eigenspace", i.e., the finite-dimensional subspace spanned by the generalized eigenforms for the U operators corresponding to eigenvalues with valuation $\alpha$. As we have seen, $e_\alpha$ is a continuous linear endomorphism of the $p$-adic Banach space $\mathsf{M}(B,k,\mathrm{N};r)\otimes K$ (which is given the $p$-adic topology for which the $B$-module $\mathsf{M}(B,k,\mathrm{N};r)$ is the closed unit ball), but it is *not* continuous on $\mathsf{M}^\dagger(B,k,\mathrm{N};1)\otimes$

$K$ when this is given the $q$-expansion topology (except when $\alpha = 0$). The first important question, then, which we have already mentioned above, is whether for fixed $r$ the norm of $e_\alpha$ is bounded independent of the weight $k$.

**Question III.2** *Let $e_\alpha^{(k)}$ denote the slope $\alpha$ projection on $\mathsf{M}(B,k,\mathrm{N};r)\otimes K$. Does there exist a bound $C(\alpha,r)$ (independent of $k$) so that $\|e_\alpha\| \le C(\alpha,r)$? If so, can we take $C(\alpha) = p^\alpha$?*

As we remarked in Chapter II, the answer to both questions is yes when $\alpha = 0$; we conjectured above (Conjecture II.4) that the bound $C(\alpha,r)$ always exists, but we have no idea whether it is plausible that $C(\alpha,r) = p^\alpha$ in general.

Closely related are the following two questions:

**Question III.3** *Suppose $f \in \mathsf{M}^\dagger(B,k,\mathrm{N};1)$ satisfies $e_\alpha f = f$; so that in particular we have $f \in \mathsf{M}(B,k,\mathrm{N};r)\otimes K$ for any $r$ such that* $\mathrm{ord}(r) < p/(p+1)$. *If $k$ is large enough, do we in fact have $f \in M(B,k,\mathrm{N})\otimes K = M(K,k,\mathrm{N})$? In other words, is it true that any overconvergent modular form of slope $\alpha$ and sufficiently high weight is necessarily a* classical *modular form?*

If $\alpha = 0$, the answer is yes, and "sufficiently high" is $k \ge 3$, as was shown by Hida (see Section II.4 above). If $f$ is an *eigenform*, we already know that $f$ must be *congruent* to a classical modular form; if we can show that this classical modular form must also be of slope $\alpha$, this would answer the question affirmatively in the case of eigenforms.

Finally, we have:

**Question III.4** *Let $P_k(t)$ denote the characteristic power series of the $U$ operator acting on $\mathsf{M}^\dagger(B,k,\mathrm{N};1)\otimes K$. Is it true that if $k_1 \equiv k_2 \pmod{p^{n-1}(p-1)}$ then $P_{k_1}(t) \equiv P_{k_2}(t) \pmod{p^n}$? If so, is the variation in fact locally analytic in $k$?*

Above, we conjectured that the answer to the first part of this question is "yes", but made a guess that the answer to the second question is "no", unless we consider truncations of the full characteristic power series at the various slopes.

Going on to questions suggested by the work in this chapter, the most natural and important question has already been stated:

**Question III.5** *Let $\overline{\rho}$ be a residual Galois representation attached to a residual eigenform $\overline{f}$. Is it true that every deformation of $\overline{f}$ is attached to a $p$-adic modular form reducing to $\overline{f}$?*

Less ambitiously, one could ask

**Question III.6** *Let $\overline{\rho}$ be a residual Galois representation attached to a residual eigenform $\overline{f}$. What is the Krull dimension of the modular deformation ring $\mathbf{R}(\overline{f})$?*

We have shown that this Krull dimension is at least three, and we have given an argument that suggests that when $\bar{f}$ is *ordinary* (i.e., of slope 0) this dimension should in fact be at least four. It is not clear to what extent this dimension will depend on $\bar{f}$ (rather than, say, only on its weight and level).

Another category of question relating to the construction of the Galois representations attached to the various deformations of a residual eigenform has to do with the relation between representation-theoretic properties of the representation deforming $\bar{\rho}$ and the properties of the modular form attached to it. For example, Mazur and Wiles have shown that any representation attached to an ordinary modular form must satisfy a certain condition on the action of an inertia group at $p$ (see [MW86]).

**Question III.7** *Let $\bar{\rho}$ be an absolutely irreducible residual representation attached to a residual eigenform $\bar{f}$. Let $\rho$ be any modular deformation, and let $f$ be the $p$-adic modular form attached to it. Can one find a condition on $\rho$ that will hold if and only if $f$ is overconvergent? Is it true that $f$ is ordinary if and only if $\rho$ satisfies the condition of Mazur and Wiles?*

There are several other crucial questions about the modular deformation space which have been touched upon only lightly in this chapter; they have to do with the dependence of our construction on the level. As we pointed out above, we have always assumed that the level N of our modular forms was fixed beforehand, and that all modular forms under consideration were ($p$-adically) of level N. (Recall that this includes classical modular forms of level $\mathrm{N}p^{\nu}$ for every $\nu$; in general, we may always assume that all levels are prime to $p$, since introducing powers of $p$ does not alter the spaces of $p$-adic modular forms in question.) In fact, there are two possible ways to vary the level: adding new prime divisors to the level or not.

To begin with, let $\mathrm{N}_1$ be a number prime to $p$, and let N be the product of its prime divisors (this is sometimes called the *radical* of $\mathrm{N}_1$). It is clear that we have an inclusion $\mathbf{V}(B,\mathrm{N}) \hookrightarrow \mathbf{V}(B,\mathrm{N}_1)$, and hence that we have an epimorphism of Hecke algebras $\mathbf{T}(B,\mathrm{N}_1) \longrightarrow \mathbf{T}(B,\mathrm{N})$; localizing and completing at the maximal ideal corresponding to some eigenform $\bar{f} \in \mathbf{V}(\mathsf{k},\mathrm{N})$, we get a map between the modular deformation ring of level $\mathrm{N}_1$ and the modular deformation ring of level N:

$$\mathbf{R}^{(\mathrm{N}_1)}(\bar{f}) \longrightarrow \mathbf{R}^{(\mathrm{N})}(\bar{f}).$$

Note that these are both deformation rings for the representation associated to $\bar{f}$, since "unramified outside $\mathrm{N}p$" and "unramified outside $\mathrm{N}_1p$" are synonymous.

**Question III.8** *What is the relation between these two rings of deformations outside $\mathrm{N}p$? Do they have the same Krull dimension? Can they be distinguished representation-theoretically, say, by a finer examination of the ramification at the primes dividing* N*?*

We may, of course, take the inverse limit over all $\mathrm{N}_1$ with the same prime divisors, and get a deformation of the representation associated to $\bar{f}$ to an even larger ring

$$\mathbf{R}^{(\infty)}(\bar{f}) = \varprojlim_{\mathrm{N}_1} \mathbf{R}^{(\mathrm{N}_1)}(\bar{f}).$$

**Question III.9** *What are the properties of this larger deformation ring* $\mathbf{R}^{(\infty)}(\bar{f})$*? In particular, what is its Krull dimension? Is it equal to Mazur's full deformation ring? Is it equal to the modular deformation ring of level* $\mathrm{N}_1$ *if* $\mathrm{N}_1$ *is sufficiently large?*

In the case where N|M and there are primes dividing M which do not divide M, the situation is even more interesting. If we consider a residual eigenform $\bar{f} \in \mathbf{V}(B, \mathrm{N})$, we may think of it as being of level N, in which case we will look for representations unramified outside N$p$, *or* of level M, in which case we will look for representations unramified outside M$p$, which form a larger space. Hence we get a map of deformation spaces

$$\mathbf{R}^{(\mathrm{M})}(\bar{f}) \longrightarrow \mathbf{R}^{(\mathrm{N})}(\bar{f}),$$

which we may think of as defining the space of deformations unramified outside N$p$ as a subscheme of the space of deformations unramified outside M$p$.

**Question III.10** *What is the relation between the spaces* $\mathbf{R}^{(\mathrm{M})}(\bar{f})$ *and* $\mathbf{R}^{(\mathrm{N})}(\bar{f})$*? Do they have the same Krull dimension?*

Even more interesting is the following situation: suppose $\bar{\rho}$ is a residual Galois representation unramified outside N$p$ and which is known to be attached to a residual eigenform of level M$p$, for some M as above. Then we can consider three deformation spaces: first, Mazur's deformation ring $\mathcal{R}^{(\mathrm{M})}(\bar{\rho})$ of $\bar{\rho}$ *considered as unramified outside* M$p$ (so we look for deformations deforming $\bar{\rho}$ which are unramified outside M$p$), second, the ring $\mathbf{R}^{(\mathrm{M})}(\bar{\rho})$, corresponding to (level M) *modular* deformations, and third, the ring $\mathcal{R}^{(\mathrm{N})}(\bar{\rho})$, corresponding to deformations (modular or not) which are unramified outside N$p$. We have a diagram of surjections:

$$\begin{array}{ccc} & & \mathcal{R}^{(\mathrm{N})}(\bar{\rho}) \\ & \nearrow & \\ \mathcal{R}^{(\mathrm{M})}(\bar{\rho}) & & \\ & \searrow & \\ & & \mathbf{R}^{(\mathrm{M})}(\bar{\rho}) \end{array}$$

We can interpret this as defining two subspaces of the space of deformations unramified outside M$p$: the subspace of level M modular deformations and the subspace of deformations unramified outside N$p$.

**Question III.11** *What is the intersection of these two subspaces? In particular, is it always non-empty? In other words, does the existence of deformations unramified outside* N$p$ *imply the existence of* modular *deformations which are of level* N*?*

Making N precise is of course crucial here. If this is done correctly, an affirmative answer is of course expected, since the problem can be reworded as a part of a well-known conjecture due to Serre:

**Conjecture III.1** *Suppose $\overline{\rho}$ is a residual representation attached to a residual eigenform $\bar{f}$ of (weight $k$ and) squarefree level* M*, and suppose that $\overline{\rho}$ is in fact unramified at some prime $\ell$ dividing* M*. Then there exists a residual eigenform of (possibly different weight and) level* $\mathrm{N} = \mathrm{M}/\ell$ *to which $\overline{\rho}$ is attached. Put in other words, if $\bar{f}$ is the reduction of a classical modular form of level* M *whose attached representation is unramified outside* $\mathrm{M}/\ell$*, then there exists a classical modular form of level* $\mathrm{M}/\ell$ *(but possibly of different weight) which also reduces to $\bar{f}$.*

Serre's conjecture, of course, goes on to define precisely the minimal possible weight for the lifting. It seems to us that the "level part" of the conjecture should be accessible by $p$-adic methods (the "weight part" is probably not). One might attempt, for example, to prove it first in the ordinary case.

# Bibliography


[AS86] Avner Ash and Glenn Stevens. "Cohomology of Arithmetic Groups and Congruences Between Systems of Hecke Eigenvalues". *Journ. Reine Angew. Math.*, 365:192–220, 1986.

[CPS] E. Cline, B. Parshall, and L. Scott. "Cohomology of Finite Groups of Lie Type, I". *Publ. Math. I.H.E.S.*, 45:169–215, 1975.

[De71] P. Deligne. "Formes Modulaires et Représentations $\ell$-adiques", exposé 355. In *Séminaire Bourbaki 1968/1969*, pages 139–172, Springer-Verlag, 1971.

[DeRa] P. Deligne and M. Rapoport. "Les Schémas de Modules de Courbes Elliptiques". In *Modular Forms in One Variable II (SLN 349)*, Springer-Verlag, Berlin, Heidelberg, New York, 1973.

[DS74] P. Deligne and J. P. Serre. "Formes Modulaires de Poids 1". *Ann. Sci. E. N. S.*, 7:507–530, 1974. In J. P. Serre, *Œuvres*, volume III.

[Dw73] B. Dwork. "The $U_p$ Operator of Atkin on Modular Functions of Level 2 with Growth Conditions". In W. Kuijk and Jean-Pierre Serre, editors, *Modular Forms in One Variable III (SLN 350)*, Springer-Verlag, Berlin, Heidelberg, New York, 1973.

[EGA] A. Grothendieck and J. Dieudonné. *Eléments de Géométrie Algébrique*:

- 0: "Préliminaires", distributed through chapters I–IV.
- I: "Le langage des schémas", *Publ. Math. I. H. E. S.*, 4 (1960); revised as *Eléments de Géométrie Algébrique I*, Springer-Verlag, Berlin, Heidelberg, New York, 1971.
- II: "Étude globale élémentaire de quelques classes de morphismes," *Publ. Math. I. H. E. S.*, 8 (1961).
- III: "Étude cohomologique des faisceaux cohérents," *Publ. Math. I. H. E. S.*, 11 (1961), and 17 (1963).
- IV: "Étude locale des schémas et des morphismes des schémas," *Publ. Math. I. H. E. S.*, 20 (1964), 24 (1965), 28 (1966), and 32 (1967).

[Hi] H. Hida, letter to the author.

[Hi86a] H. Hida. "Galois Representations into $GL_2(\mathbf{Z}_p[[X]])$ Attached to Ordinary Cusp Forms". *Inv. Math.*, 85:545–613, 1986.

[Hi86b] H. Hida. "Iwasawa Modules Attached to Congruences of Cusp Forms". *Ann. Sci. Éc. Norm. Sup.*, 19:231–273, 1986.

[Jo82a] Naomi Jochnowitz. "A Study of the Local Components of the Hecke Algebra mod $\ell$". *Trans. Amer. Math. Soc.*, 270:253–267, 1982.

[Jo82b] Naomi Jochnowitz. "Congruences Between Systems of Eigenvalues of Modular Forms". *Trans. Amer. Math. Soc.*, 270:269–285, 1982.

[Ka73] Nicholas M. Katz. "$p$-adic Properties of Modular Schemes and Modular Forms". In W. Kuijk and Jean-Pierre Serre, editors, *Modular Forms in One Variable III (SLN 350)*, Springer-Verlag, Berlin, Heidelberg, New York, 1973.

[Ka75a] Nicholas M. Katz. "Higher Congruences Between Modular Forms". *Ann. of Math.*, 101:332–367, 1975.

[Ka75b] Nicholas M. Katz. "$p$-adic L-functions via Moduli of Elliptic Curves". In Robin Hartshorne, editor, *Algebraic Geometry: Arcata 1974*, American Mathematical Society, Providence, Rhode Island, 1975.

[Ka76] Nicholas M. Katz. "$p$-adic Interpolation of Real Analytic Eisenstein Series". *Ann. of Math.*, 104:459–571, 1976.

[Ka77] Nicholas M. Katz. "The Eisenstein Measure and $p$-adic Interpolation". *Am. Journ. Math.*, 99:238–311, 1977.

[KM85] Nicholas M. Katz and Barry Mazur. *Arithmetic Moduli of Elliptic Curves*. Volume 108 of *Annals of Mathematics Studies*, Princeton University Press, Princeton, New Jersey, 1985.

[Ko84] Neal Koblitz. *Introduction to Elliptic Curves and Modular Forms*. Springer-Verlag, Berlin, Heidelberg, New York, 1984.

[KS73] W. Kuijk and J-P. Serre, editors. *Modular Funtions of One Variable III*. Volume 350 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, Heidelberg, New York, 1973.

[La76] Serge Lang. *Introduction to Modular Forms*. Springer-Verlag, Berlin, Heidelberg, New York, 1976.

[Ma] Barry Mazur. "Deforming Galois Representations". preprint.

[MTT86] B. Mazur, J. Tate, and J. Teitelbaum. "On $p$-adic Analogues of the Conjectures of Birch and Swinnerton-Dyer". *Inv. Math.*, 84:1–48, 1986.

[MW86] B. Mazur and A. Wiles. "On $p$-adic Analytic Families of Galois Representations". *Comp. Math.*, 59:231–264, 1986.

[Monn70] A. F. Monna. *Analyse Non-Archimedienne.* Springer-Verlag, Berlin, Heidelberg, New York, 1970.

[Mons71] Paul Monsky. "Formal Cohomology: III. Fixed Point Theorems". *Ann. of Math.(2)*, 93:315–343, 1971.

[MoWa68] Paul Monsky and G. Washnitzer. "Formal Cohomology I". *Ann. of Math.(2)*, 88:181–217, 1968.

[Ogg69] Andrew Ogg. *Modular Forms and Dirichlet Series.* W. A. Benjamin, New York, 1969.

[Oht82] Ohta. "On $\ell$-adic Representations attached to Automorphic Forms". *Jap. Journ. Math.*, 8 (new series):1–47, 1982.

[Sch68] M. Schlessinger. "Functors of Artin Rings". *Trans. A. M. S.*, 130:208–222, 1968.

[Se62] Jean-Pierre Serre. "Endomorphismes Complètement Continus des Espaces de Banach $p$-adiques". *Publ. Math. I.H.E.S.*, 12:69–85, 1962.

[Se73] Jean-Pierre Serre. "Formes Modulaires et Fonctions Zêta $p$-adiques". In W. Kuijk and Jean-Pierre Serre, editors, *Modular Forms in One Variable III (SLN 350)*, Springer-Verlag, Berlin, Heidelberg, New York, 1973.

[Shi68] Goro Shimura. "An $\ell$-adic Method in the Theory of Automorphic Forms". 1968. unpublished.

[Shi71] G. Shimura. *Introduction to the Arithmetic Theory of Automorphic Forms.* Princeton University Press, 1971.

[Ta67] J. T. Tate. "$p$-divisible Groups". In T. A. Springer, editor, *Proceedings of a Conference on Local Fields*, pages 158–183, Springer-Verlag, Berlin, Heidelberg, New York, 1967.

# Index