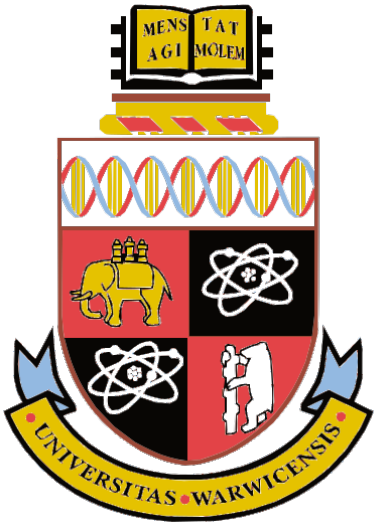


Slopes of p-adic modular forms

by

André Macedo

MASt Thesis

Supervised by David Loeffler

Submitted to The University of Warwick

Mathematics Institute

April, 2016

Contents

1 Introduction

Given a prime $p \geq 5$ and a positive integer N (coprime to p), the Atkin operator U_p acts linearly on the space of classical modular forms of weight k and level $\Gamma_1(N) \cap \Gamma_0(p)$. The eigenvalues of this map are extremely important in Number Theory since they encode arithmetic information about modular forms and their Galois representations. Moreover, if f is an eigenform for U_p with eigenvalue λ, the p-adic valuation (normalized so that $v_p(p) = 1$) of λ, also called the *slope*[1] of f, plays an essential role in the p-adic theory of modular forms. Given a rational number α, we let $\mathbf{d}(k, \alpha)$ denote the number of U_p-eigenvalues with p-adic valuation α. In the 1990's, Gouvêa and Mazur conjectured that this function exhibits a striking p-adic continuity on the variable k. Although this conjecture was recently proved to be false, a weaker form of continuity is true. In fact, Wan (extending results of Coleman) proved that if k_1 and k_2 are two integers (both at least $2\alpha + 2$) which are congruent modulo $p^n(p-1)$ for some integer $n = O(\alpha^2)$, then $\mathbf{d}(k_1, \alpha) = \mathbf{d}(k_2, \alpha)$.

The main goal of this project is to present Wan's proof of this result. This proof will require us to use various tools from p-adic analysis (like Newton polygons and the Fredholm determinant) and to introduce spaces of p-adic modular forms. The theory of p-adic modular forms was founded in the 1970's by Serre, who defined a p-adic modular form as the p-adic limit of some compatible family of q-expansions of classical modular forms. However, it can be shown that this space contains U_p-eigenforms with arbitrary eigenvalue, so these p-adic eigenfunctions do not contain any interesting arithmetical information. This was corrected by Katz, which soon afterwards introduced a notion of p-adic modular form which generalized Serre's construction to a much more geometrical context. Katz's insight is that we should define these forms in modular terms, i.e., as functions on elliptic curves with some additional structure. This is achieved by considering the rigid analytic space obtained by removing p-adic discs around the supersingular points in the compactified space $X_1(N)$ of elliptic curves with a $\Gamma_1(N)$-structure. The concept of an overconvergent modular form will then arise as a rule on this rigid analytic space satisfying some transformation properties. This definition will give us a space that extends the space of classical modular forms and for which the U_p operator is compact, so we have a good spectral theory. As we will see, this space is extremely useful to study slopes. In fact, if k is greater than $\alpha + 1$, the number of generalized eigenforms of weight k and slope α is the same for both classical and overconvergent modular forms. This remarkable connection between slopes of classical and overconvergent modular forms will then allow us to construct the quadratic bound that figures in Wan's theorem.

[1] This name comes from the theory of Newton polygons, which tells us that these p-adic valuations are determined by the slopes of the Newton polygon of the characteristic polynomial of U_p. This will be further explained in Chapter 2.

2 Newton polygons

In the area of p-adic analysis, the Newton polygon is an important tool that allows us to understand the behaviour of polynomials (or, more generally, p-dic power series) over local fields. We will mainly work with the field $\mathbb{C}_p$, the completion of the algebraic closure of the p-adic field $\mathbb{Q}_p$. This is the most natural field to work with if we want a good theory of p-adic analysis, as it is both complete and algebraically closed.

2.1 Newton polygons for polynomials

Let $f(x) \in \mathbb{C}_p[x]$ be a polynomial. Since we are mainly interested in understanding how the zeros of f behave, we may suppose (factoring any powers of x if necessary) that $f(0) \neq 0$. Also, dividing through by $f(0)$, we can assume that $f(0) = 1$ and so we are now looking at a polynomial of the form $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots \ a_1 x + 1$, with $a_i \in \mathbb{C}_p$ for $1 \leq i \leq n$ and $a_n \neq 0$. To define the Newton polygon of f, we first need to plot the set of points

$$X_f = \{(0,0), (1, v_p(a_1)), \cdots, (n-1, v_p(a_{n-1})), (n, v_p(a_n))\}$$

in the plane $\mathbb{R}^2$, where v_p extends[2] the usual p-adic valuation to $\mathbb{C}_p$, normalized so that $v_p(p) = 1$. Note that if $a_i = 0$ for some i, the quantity $v_p(a_i)$ is ∞, so we just skip this point in the definition of X_f.

Definition 2.1. *The Newton polygon of $f(x)$ is defined to be the "lower boundary" of the convex hull of X_f in $\mathbb{R}^2$, i.e., the longest convex polygonal line joining the first point $(0,0)$ with the last point $(n, v_p(a_n))$ which passes on or below all of the points in X_f.*

Geometrically, the construction of the Newton polygon is very easy to explain: start with the vertical line $x = 0$ and rotate it counterclockwise until it hits one of the points in X_f. When this happens, just "break" the line at that point (if the line hits various points at the same time, this break is made at the last such point), and keep rotating the remaining part of the line until a new point is hit. If we continue this process until all the points in X_f have either been hit or are located strictly above the polygonal line, we get the Newton polygon of $f(x)$ (in practice, when this process ends, we cut off the remaining infinitely long line at the last vertex).

In order to see how the Newton polygon allows us to extract information about the roots of a polynomial, we need to define two more concepts:

- The slopes of the line segments in the Newton polygon are called the *slopes of the Newton polygon of f* or simply the *slopes of f*;

[2] For more information on extension of norms, see [13, pg. 57].

- The *length* of each slope is the length of the projection of the corresponding line segment on the x-axis.

The first interesting feature of the Newton polygon of a polynomial f is that, in a very precise sense, its slopes are "recording" the p-adic valuation of the roots of f:

Proposition 2.2. *Let $f(x) = c(x - \alpha_1) \cdots (x - \alpha_n)$ be a polynomial in $\mathbb{C}_p[x]$ and λ a slope of its Newton polygon with corresponding lenght L. Define also $\lambda_i = v_p(\frac{1}{\alpha_i})$, for all $1 \leq i \leq n$. Then, exactly L of the λ_i (counting with multiplicity) are equal to λ.*

Proof. See Koblitz [13, IV, p. 98]. □

In other words, this proposition tells us that f has exactly L roots (counting multiplicities) of p-adic norm p^{λ}. This is quite useful since it allows us to quickly identify at what radii the roots of f are located.

2.2 Newton polygons for power series

As we have seen in the previous section, Newton polygons codify information concerning the factorization of polynomials. With this fact in mind, it is very natural to take the next step and consider Newton polygons of a power series.

Definition 2.3. *Let $f(x) = 1 + \sum_{i=1}^{\infty} a_i x^i$ be a power series with coefficients in $\mathbb{C}_p$ and $f_n(x) = 1 + \sum_{i=1}^{n} a_i x^i$ be the n-th partial sum of $f(x)$. We define the Newton polygon of f as the limit of the Newton polygons of $f_n(x)$, i.e., the polygon obtained as the limit of the "rotating line" procedure applied to X_{f_n}.*

In general, we can use the geometric algorithm of rotating the line $x = 0$ in order to draw the Newton polygon of a power series (with the same definition for X_f), but there are some special cases that should be noted:

- It can happen that in some step of the procedure the line simultaneously hits infinitely many of the points we plotted, e.g., the power series $f(x) = 1 + \sum_{i=1}^{\infty} x^i$. If this is the case, we stop at that point (with this last segment being infinitely long) and the Newton polygon is then complete.
- The polygonal line reaches a position where it cannot be rotated further without missing some points in X_f. For example, the Newton polygon of $f(x) = 1 + \sum_{i=1}^{\infty} px^i$ has this problem when the rotating line reaches the horizontal position. In this case, we also stop at this step (with the last segment having slope equal to the supremum of all possible slopes for which the line passes bellow all of the points in X_f) and the Newton polygon is complete.

The Newton polygon of a power series also contains a lot of information about the location of its zeros and even its convergence. We present three results that illustrate this fact quite well and that will prove to be useful in a later section:

Proposition 2.4. *Let $f(x) = 1 + \sum_{i=1}^{\infty} a_i x^i \in \mathbb{C}_p[[x]]$ and denote $m = \sup\{\lambda : \lambda \text{ is a slope of } f\}$. Then the radius of convergence[3] of the series is p^m (which should be interpreted as ∞, if $m = \infty$).*

Proof. The bulk of the proof is to use the nice fact that, in the p-adic metric, a power series $f(x) = 1 + \sum_{i=1}^{\infty} a_i x^i$ converges at $x = x_0$ if and only if the p-adic norm $\left\|a_i {x_0}^i\right\|_p \to 0$ (or equivalently that $v_p(a_i {x_0}^i) \to \infty$) as $i \to \infty$.

Now, suppose that $\|x\|_p < p^m$, i.e., $-v_p(x) = b < m$. Using the fact of the previous paragraph, in order to show that f converges at x we need to show that $v_p(a_i x^i) = v_p(a_i) - ib \to \infty$. But since $m > b$, it is clear that, if we overlap the line $L : y = bx$ with the Newton polygon of f, the polygon gets farther and farther above L so that the plotted points $(i, v_p(a_i))$ also get farther and farther above L. Hence, it follows that $v_p(a_i) - ib \to \infty$ as $i \to \infty$, as desired. The proof that $\|x\|_p > p^m$ implies that f does not converge at x uses the exact same idea, so we omit it. □

Theorem 2.5 (p-adic Weierstrass Preparation Theorem)**.** *Let K be an extension of $\mathbb{Q}_p$ which is complete and write $\mathcal{O}$ for the valuation ring $\mathcal{O} = \{x \in K : \|x\|_p \leq 1\}$. Let $f(x) = \sum_{i=0}^{\infty} a_i x^i \in K[[x]]$ be a power series such that $a_n \to 0$ as $n \to \infty$, so that $f(x)$ converges for all points in $\{x \in \mathbb{C}_p : \|x\|_p \leq 1\}$. Let N be the number defined by the conditions*

$$\|a_N\|_p = \max\{\|a_n\|_p : n \geq 0\} \text{ and } \|a_n\|_p < \|a_N\|_p \text{ for all } n > N.$$

Then there exists a polynomial

$$g(x) = b_0 + b_1 x + \cdots + b_N x^N \in K[x]$$

of degree N and a power series

$$h(x) = 1 + \sum_{i=1}^{\infty} c_i x^i \in K[[x]]$$

satisfying

1. $f(x) = g(x)h(x)$
2. $\|b_N\|_p = \max_n \|b_n\|_p$, *i.e.*, $\|g(x)\|_1 = \|b_N\|_p$
3. $\|c_n\|_p < 1$ *for all* $n \geq 1$, *i.e.*, $\|h(x) - 1\|_1 < 1$
4. $\|f(x) - g(x)\|_1 < 1$.

[3] As in the case of real analysis, given a power series $f(x) = \sum_{i=0}^{\infty} a_i x^i \in \mathbb{C}_p[[x]]$, the radius of convergence r of f is a non-negative real number or ∞ (with the obvious interpretation) such that f converges in $\{x \in \mathbb{C}_p : \|x\|_p < r\}$ and diverges in $\{x \in \mathbb{C}_p : \|x\|_p > r\}$.

In particular, $h(x)$ has no zeros in $\mathcal{O}$.

Proof. See Gouvêa [7, VI, Theorem 6.3.6]. □

Corollary 2.6. *Let $f(x) = 1 + \sum_{i=1}^{\infty} a_i x^i \in \mathbb{C}_p[[x]]$ be an entire p-adic power series, i.e., a power series which converges in all of $\mathbb{C}_p$. Then $f(x)$ can be written as an infinite product*

$$f(x) = \prod_{i=1}^{\infty} (1 - \lambda_i x)$$

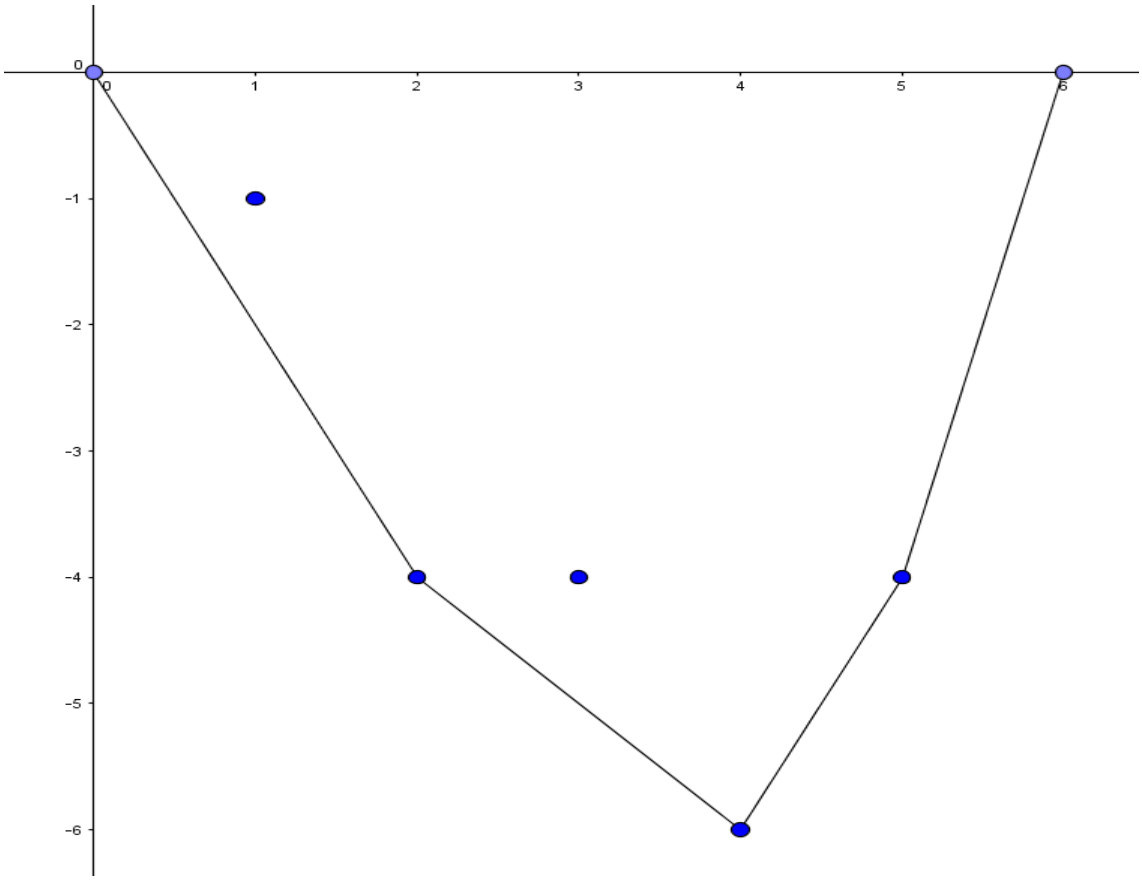
where λ_i are the reciprocals of the zeros of $f(x)$.

Proof. See Gouvêa [7, VI, Proposition 6.4.1]. □

2.3 Examples

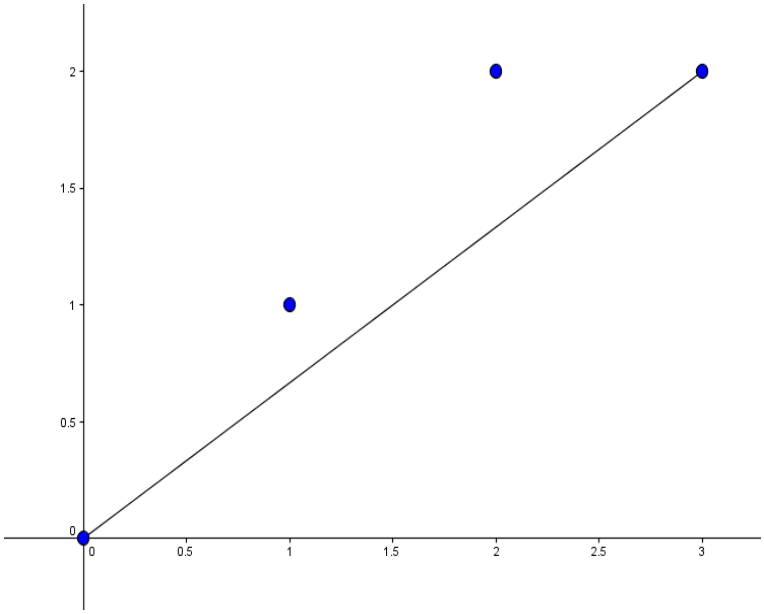
In this section we present some examples of Newton polygons and give an interesting irreducibility criterion based on them.

Example 2.7. Let $f(x) = x^6 - \frac{197}{16}x^5 + \frac{3569}{64}x^4 - \frac{1799}{16}x^3 + \frac{1517}{16}x^2 - \frac{43}{2}x + 1 \in \mathbb{C}_2[x]$. Then, we have $X_f = \{(0,0), (1,-1), (2,-4), (3,-4), (4,-6), (5,-4), (6,0)\}$ so that the Newton polygon of $f(x)$ becomes:



Using Proposition 2.2, this picture tells us immediately that f has two roots with 2-adic valuation 2, two roots with 2-adic valuation 1, one root with 2-adic valuation -2 and one root with 2-adic valuation -4. In fact, the roots of f are 4,2 (both with multiplicity 2) and $\frac{1}{4}, \frac{1}{16}$ (both with multiplicity 1), which confirms the previous 2-adic analysis.

Example 2.8. Let $f(x) = 1 + 6x + 9x^2 + 18x^3 \in \mathbb{C}_3[x]$. Start by noticing that $f(x)$ is irreducible over $\mathbb{Z}$, since it has degree 3 and no integer roots. Moreover, we have $X_f = \{(0,0),(1,1),(2,2),(3,2)\}$, so that the Newton polygon of $f(x)$ becomes just a straight line segment:



This example illustrates the following fact: if the Newton polygon of a polynomial $f(x) \in 1 + x\mathbb{Z}_p[x]$ of degree m consists of one line segment joining $(0,0)$ to the point (m,n) where $\gcd(m,n) = 1$, then $f(x)$ is irreducible over $\mathbb{Z}_p$. This is because by Proposition 2.2 any root α of $f(x)$ has p-adic valuation $-\frac{n}{m}$. But if α satisfies a polynomial of degree $d < m$, then it is clear that $v_p(\alpha)$ must also be a fraction with denominator at most d since $v_p(\alpha)$ is the symmetric of a slope in a Newton polygon of a polynomial with degree d. Since $v_p(\alpha) = -\frac{n}{m}$ is in irreducible form, we must have $m \leq d$, a contradiction.

This fact can also be used to give a quick proof of Eisenstein's criterion. In fact, if $p(x) = a_0 + a_1x + \cdots + a_nx^n \in \mathbb{Z}[x]$ is a polynomial of degree n which is Eisenstein at a prime p, then we know that

1. $v_p(a_0) = 1$
2. $v_p(a_i) \geq 1$ for $0 < i < n$
3. $v_p(a_n) = 0$.

Now, dividing $p(x)$ by a_0, we get a polynomial $q(x) = 1 + b_1x + \cdots + b_nx^n \in \mathbb{Z}[x]$ which satisfies

1. $q(0) = 1$
2. $v_p(b_i) \geq 0$ for $0 < i < n$
3. $v_p(b_n) = -1$.

From these three conditions it is clear that the Newton polygon of $q(x)$ is the line segment joining $(0,0)$ to $(n,-1)$, so by the previous criterion $q(x)$ is irreducible, which implies that $p(x)$ is also irreducible.

3 p-adic modular forms

In this chapter we introduce the p-adic theory of classical modular forms and overconvergent modular forms. This is the theory necessary to work with the main problem studied in this project, namely the Gouvêa-Mazur conjecture. Throughout the section, we will fix a prime number $p \geq 5$ and $N \in \mathbb{Z}_{>0}$ coprime to p. We will also work with a finite field extension K of the p-adic numbers $\mathbb{Q}_p$ and we will let B denote the ring of integers of K.

3.1 Classical modular forms

The main object of focus in this section is the space of classical modular forms of weight $k \in \mathbb{Z}_{\geq 0}$ and level $\Gamma_1(N) \cap \Gamma_0(p)$. As is well-known (see Diamond and Shurman [6, III]), this is a finite dimensional vector space over $\mathbb{C}$ with a basis β consisting of modular forms whose q-expansions at infinity have rational coefficients. This leads us to our first definition:

Definition 3.1. *We denote by $M_k(Np, B)$ the free B-module generated by β.*

The Atkin operator U_p plays a very significant role in studying the structure of this space. This operator acts on $M_k(Np, B)$ and it is defined by its effect on the q-expansion of $f = \sum_{n=0}^{\infty} a_n q^n \in M_k(Np, B)$:

$$U_p\left(\sum_{n=0}^{\infty} a_n q^n\right) = \sum_{n=0}^{\infty} a_{np} q^n.$$

We also introduce the characteristic polynomial of U_p acting on $M_k(Np, B)$, $\mathbf{P}_k(t) = \det(1 - tU_p)$. This is a polynomial in $B[t]$ and, for our purposes, it is useful to consider the factorization of $\mathbf{P}_k(t)$ into its characteristic roots (reciprocal of the eigenvalues)

$$\mathbf{P}_k(t) = \prod_{\alpha \in \mathbb{Q}} \mathbf{P}_k(t)^{(\alpha)}$$

where $\mathbf{P}_k(t)^{(\alpha)}$ is the factor of $\mathbf{P}_k(t)$ consisting of the characteristic roots $\boldsymbol{\rho}_{\boldsymbol{\alpha}}$ of $\mathbf{P}_k(t)$ with p-adic valuation[4] $v_p(\boldsymbol{\rho}_{\boldsymbol{\alpha}}) = -\alpha$. This decomposition of $\mathbf{P}_k(t)$ leads us to our next definition:

Definition 3.2. *We denote the degree of $\boldsymbol{P}_k(t)^{(\alpha)}$ by $\mathbf{d}(k, \alpha)$, i.e., $\mathbf{d}(k, \alpha)$ is the number (counting with multiplicity) of U_p-eigenvalues with p-adic valuation α.*

Example 3.3. Let $N = 1$, $p = 3$ and $k = 4$. Then we are looking at $M_4(3, B)$, the space of modular forms of weight 4 and level $\Gamma_0(3)$. This space is 2-dimensional with basis[5] $\beta = \{f_1, f_2\}$, where $f_1 = 1 + 240q^3 + 2160q^6 + O(q^9)$ and $f_2 = q + 9q^2 + 27q^3 + 73q^4 +$

[4] Hereinafter, we always assume that the p-adic valuation v_p is normalized so that $v_p(p) = 1$.

[5] I used the SageMath software to perform these calculations.

$126q^5 + 243q^6 + O(q^7)$. Hence, a simple calculation shows that the U_3-eigenvalues on this space are 1 and 27, and so we have $\mathbf{d}(4,3) = \mathbf{d}(4,0) = 1$ and $\mathbf{d}(4,\alpha) = 0$ for $\alpha \neq 0,3$.

An interesting object of study is the variation of the dimension $\mathbf{d}(k,\alpha)$ with the variation of the weight k (where the slope[6] α is fixed). In fact, the investigation of this problem is related to the existence of p-adic families of eigenforms and congruences of modular forms (see [8, I] or [9]). In 1992, Gouvêa and Mazur [9, pg. 797] conjectured the following result concerning this variation:

Conjecture 3.4 (Gouvêa-Mazur conjecture)**.** *Fix a slope α, and let k_1, k_2 be integers such that:*

- $k_1, k_2 \geq 2\alpha + 2$
- $k_1 \equiv k_2 \pmod{p^n(p-1)}$ *for some integer $n \geq \alpha$*

Then we have $\mathbf{d}(k_1,\alpha) = \mathbf{d}(k_2,\alpha)$.

This conjecture is claiming that the function $\mathbf{d}(k,\alpha)$ exhibits a remarkable p-adic continuity on the variable k. However, in 2003, Buzzard and Calegari (see [2]) presented a counterexample to this conjecture. Namely, if $p = 59$ and $N = 1$, they proved that there exists a rational number α such that $0 \leq \alpha \leq 1$ and $\mathbf{d}(16,\alpha) \neq \mathbf{d}(3438,\alpha)$. Since $3438 \equiv 16 \pmod{58 \times 59}$, this means that the above conjecture, as stated, is false. Nonetheless, Coleman proved that a weaker version of this conjecture is true, namely we have:

Theorem 3.5 (Coleman, [4])**.** *Let k_1, k_2 be two integers such that $k_1, k_2 \geq 2\alpha + 2$. Then, there exists a finite number $\mathbf{m}(\alpha)$ such that, if k_1, k_2 satisfy the congruence*

$$k_1 \equiv k_2 \pmod{p^{\mathbf{m}(\alpha)}(p-1)},$$

then $\mathbf{d}(k_1,\alpha) = \mathbf{d}(k_2,\alpha)$.

This result was soon afterwards improved by Wan, who established a quadratic bound on the quantity $\mathbf{m}(\alpha)$ of Coleman's theorem. The main objective of the rest of the project is to prove this quadratic bound. This will require us to introduce some theory of p-adic modular forms and to study the space of overconvergent modular forms, which will also support an action of the U_p operator. It turns out that the problem of counting eigenvalues with fixed p-adic valuation in this space is closely related to the problem of counting eigenvalues in the space of classical modular forms. To make this connection clear, we will need some p-adic results like the continuity of the characteristic series of U_p (to be introduced later) in the weight variable and the fact that an overconvergent U_p-eigenform with small slope is classical.

[6] It makes sense to call α a slope since $\alpha = -v_p(\boldsymbol{\rho}_{\boldsymbol{\alpha}})$, so by Proposition 2.2 α is a slope of the Newton polygon of $\mathbf{P}_k(t)$.

3.2 Overconvergent p-adic modular forms

3.2.1 The geometric viewpoint

The first attempt of defining a p-adic version of a modular form was made by Serre in [15]. His idea was to define the q-expansion of a p-adic modular form as the p-adic limit of q-expansions of classical modular forms f_i of arbitrary weight[7] k_i. We can also define an action of the Hecke operators on this space in the usual way. This definition of p-adic modular form is a very natural one, but it turns out that this space is too big to support an interesting spectral theory. In fact, for every $\lambda \in \mathbb{C}_p$ such that $v_p(\lambda) > 0$, we can construct a p-adic modular form with eigenvalue λ for the U_p operator.

This can be corrected by considering another space of p-adic modular forms, the space of overconvergent p-adic modular forms. This space has various good properties: it supports a much richer spectral theory, it contains the space of classical modular forms as a subset and its theory can be used to design algorithms that explicitly compute classical eigenforms (see [14]).

However, the definition of overconvergent modular form requires some heavy machinery. Therefore, we will give some motivation to the concept before presenting the formal definition. This motivation is not meant to be too rigorous, but rather to give the geometric viewpoint behind the formal algebraic construction. For a more rigorous approach, the reader is referred to Katz [11].

Recall that one way of defining a classical modular form f of weight k and level 1 (say, for simplicity) is to consider the associated modular curve $Y = SL_2(\mathbb{Z})\backslash\mathbb{H}$, which can be viewed as a parameter space for isomorphism classes of elliptic curves. With this setting, one can simply define f as a rule sending a pair (E,ω) (where E is an elliptic curve on Y and ω a non-vanishing differential on E) to a complex number $f(E,\omega)$, such that $f(E,\lambda\omega) = \lambda^{-k} f(E,\omega)$ (this property encodes the usual transformation property of modular forms, which in fact can be recovered by evaluating f at $(E,\omega) = (\mathbb{C}/\mathbb{Z}\tau + \mathbb{Z}, dz)$) together with some analytic properties (encoding the fact that f is supposed to be holomorphic on $\mathbb{H}$ and at ∞).

With this approach in mind, in order to define an overconvergent p-adic modular form of weight k and level $\Gamma_1(N)$, we will consider the (compactified) modular curve $X = X_1(N)$, which we view as a space parametrizing elliptic curves with some additional level structure. A very important observation due to Katz is that there are some points on X which are "bad" and where we do not want to define our modular form. These are the supersingular points, i.e., points corresponding to elliptic curves which are supersingular

[7] With this definition, it can be shown that the weight k of a p-adic modular form is a well defined element in the set $\mathbb{Z}_p \times \mathbb{Z}/(p-1)\mathbb{Z}$.

mod p. In the rigid-analytic space[8] associated to X, these supersingular points form a subspace X^{ss} isomorphic to a finite union of open discs (in the p-adic metric). However, if we consider the complement of these discs, $X^{ord} = X \setminus X^{ss}$ (the subspace of points with good ordinary, or multiplicative, reduction), and use this space in Katz's definition of overconvergent modular forms that we are about to introduce, we recover Serre's space of p-adic modular forms. Katz's insight was that we are throwing away too many elliptic curves. In fact, we should extend the definition of our modular forms a little way into these bad supersingular areas - this will exclude a lot of badly behaved modular forms and, in the end, give us a space of modular forms with a good spectral theory from the arithmetic point of view.

This process of extending the domain of our modular form should be made using some tool that allows us to measure "supersingularity" of points in X. It turns out that the right object to consider is a form which has the property of being 0 exactly at the points of X^{ss} - an example of such a function is the $\mod p$ Hasse invariant. A lift of the Hasse invariant to characteristic 0 is given by the Eisenstein series E_{p-1} and the p-adic valuation of this form will parametrize how much we want to extend a p-adic modular form into the inside of the supersingular discs. This idea may seem odd at first, but it is in fact very natural if we want a p-adic theory of modular forms that, in a certain sense, reflects congruences of q-expansions with properties of the modular forms themselves. For example, since $E_{p-1} \equiv 1 \mod p$ (for a proof of this congruence, see [12, Corollary 6.6]), the theory should imply that E_{p-1} is an invertible form. However, this would surely be a false conclusion as the value of E_{p-1} at a supersingular elliptic curve is $0 \mod p$. This leads us to introduce the following spaces: for $r \in \mathbb{Q}$ with $0 < r < 1$, we define the space $X^{\leq r} = \{x \in X : v_p(E_{p-1}(x,\omega)) \leq r\}$, where ω is a nowhere-vanishing differential (with some restrictions to be stated later) and the valuation $v_p(E_{p-1}(x,\omega))$ is independent of choice of the non-vanishing differential ω (in those same conditions). Geometrically, this space can be thought as looking like a Riemann surface with small discs removed.

We are now ready to introduce the space of overconvergent modular forms: an r-overconvergent p-adic modular form of weight k is a function defined on pairs (E,ω), where E is an elliptic curve over B and ω a non-vanishing differential on E (with some extra conditions, to be specified later) such that $v_p(E_{p-1}(E,\omega)) \leq r$, satisfying $f(E,\lambda\omega) = \lambda^{-k} f(E,\omega)$.

3.2.2 Definition

Before we introduce the formal definition of an overconvergent p-adic modular form, we need the definition of the objects where we will define our function:

[8] For more information about this rigid-analytic approach to modular forms, see Matthew Emerton's appendix in [5, pg. 377]. The reader can think of this concept as some kind of p-adic analogue of a Riemann Surface and use his geometric pictures about complex surfaces as a guide.

Definition 3.6. *Let ρ be an element in B (the ring of integers of K). A test object of level $\Gamma_1(N)$ and growth condition ρ is a quadruple $(E\backslash B,\omega,L,Y)$, where E is an elliptic curve over B, ω a non-vanishing differential on E with B coefficients which is non-zero modulo the maximal ideal of B, L is a $\Gamma_1(N)$-level structure on E and Y an element of B satisfying $Y.E_{p-1}(E,\omega) = \rho$.*

Remark 3.7. This definition gives us an algebraic way of expressing the fact that we want our form to be defined on elliptic curves in the space $X^{\leq r}$ previously presented. In fact, if $Y.E_{p-1}(E,\omega) = \rho$, using the properties of a p-adic valuation we can conclude that $v_p(E_{p-1}(E,\omega)) \leq v_p(\rho) =: r$, so $X^{\leq r}$ will in fact be the source of our form.

Definition 3.8. *Let $k \in \mathbb{Z}$ be an integer and $\rho \in B$ be an element which is not a unit[9] in B. A ρ-overconvergent p-adic modular form of weight k, level $\Gamma_1(N)$ and growth condition ρ is a rule f sending a test object $(E\backslash B,\omega,L,Y)$ of level $\Gamma_1(N)$ and growth condition ρ to an element $f(E\backslash B,\omega,L,Y) \in B$ satisfying the following conditions:*

- *$f(E\backslash B,\lambda\omega,L,\lambda^{p-1}Y) = \lambda^{-k} f(E\backslash B,\omega,L,Y)$ for every $\lambda \in B^\times$,*
- *$f(E\backslash B,\omega,L,Y)$ depends only on the isomorphism class of the quadruple $(E\backslash B,\omega,L,Y)$,*

together with an extra condition[10] coming from the evaluation of f at the Tate elliptic curve that we omit. We also denote the space of ρ-overconvergent p-adic modular forms of weight k and level $\Gamma_1(N)$ by $M_k(N,B,\rho)$.

These are not scary objects! Indeed, as in the classical case, overconvergent modular forms have q-expansions that are a lot easier to work with. In fact, in the next chapter we will mainly work with these modular forms via the Katz's expansion (to be introduced) and refer the reader to the literature when the results we need use the previous formal definition.

3.2.3 Useful results

We now give a much more interesting description of these overconvergent modular forms. This description is given by a certain expansion, which surprisingly uses the space $M_k(N,B)$ of classical modular forms of weight k, level $\Gamma_1(N)$ and with q-expansions in $B[[q]]$. But first we need to know more about the structure of $M_k(N,B)$. Notice that, in this space, multiplication by E_{p-1} gives a subspace of $M_{k+(p-1)}(N,B)$, since multiplying two modular forms of weights w_1, w_2 produces a modular form of weight $w_1 + w_2$. It turns out that this subspace admits a complement:

[9] In fact, it is a theorem of Katz (see [11, Prop. 2.7.2]) that if we choose ρ to be a B-unit, then we would end up with Serre's space of p-adic modular forms.

[10] This condition encodes the usual analytic properties of a modular form. For more details on the nature of this condition, see [8, pg. 6].

Theorem 3.9. *Let j be a positive integer and consider the inclusion $E_{p-1}.M_{k+(j-1)(p-1)}(N,B) \subset M_{k+j(p-1)}(N,B)$. Then there is a free B-submodule $W_j(N,B)$ of $M_{k+j(p-1)}(N,B)$ such that*

$$M_{k+j(p-1)}(N,B) = E_{p-1}.M_{k+(j-1)(p-1)}(N,B) \oplus W_j(N,B)$$

where we set $W_0(N,B) = M_k(N,B)$.

Proof. See Katz [11, II, Lemma 2.6.1]. □

Let us illustrate the previous theorem with an example:

Example 3.10. Take $N = 1$, $p = 5$ and set, for all $j \geq 1$, $W_j(1,B) = < E_6^b|_{6b=k+4j}^{b\in\mathbb{Z}_{\geq 0}} >_B$ (if such a b does not exist, simply put $W_j(1,B) = \{0\}$) which is a free B-submodule of $M_{k+4j}(1,B)$. Moreover, since the B-module $M_{k+4(j-1)}$ is given by $< E_4^a E_6^b|_{4a+6b=k+4(j-1)}^{a,b\in\mathbb{Z}_{\geq 0}} >_B$, we have

$$E_4 M_{k+4(j-1)} \oplus W_j(N,B) =< E_4^{a+1} E_6^b|_{4a+6b=k+4(j-1)}^{a,b\in\mathbb{Z}_{\geq 0}} >_B \oplus < E_6^b|_{6b=k+4j}^{b\in\mathbb{Z}_{\geq 0}} >_B$$

which is clearly the space $M_{k+4j}(1,B) = < E_4^a E_6^b|_{4a+6b=k+4j}^{a,b\in\mathbb{Z}_{\geq 0}} >_B$, as desired.

We are now in perfect conditions to introduce the Katz expansion of an overconvergent modular form, using the subspaces $W_j(N,B)$:

Theorem 3.11. *Let $f \in M_k(N,B,\rho)$ be a ρ-overconvergent p-adic modular form. Then f admits the following unique expansion (called the Katz expansion of f)*

$$f = \sum_{j=0}^{\infty} \rho^j \frac{b_j}{E_{p-1}^j} \tag{1}$$

where $b_j \in W_j(N,B)$ and $\lim_{j\to\infty} b_j = 0$ (this limit just means that the q-expansions of b_j become more and more divisible by p as $j \to \infty$). Conversely, every element of the form (1) *is in $M_k(N,B,\rho)$.*

Proof. See Katz [11, II, Lemma 2.6.2]. □

Corollary 3.12. *We have the inclusion $M_k(N,B) \subset M_k(N,B,\rho)$, i.e., a classical modular form f is a ρ-overconvergent p-adic modular form.*

Proof. This follows immediately from the converse part of the theorem, putting $b_0 = f \in M_k(N,B) = W_0(N,B)$ and $b_j = 0$ for $j \geq 1$. □

3.3 The U_p operator

We also have the usual U_p operator on spaces of overconvergent modular forms, acting as expected on q-expansions:

$$U_p(\sum_{n=0}^{\infty} a_n q^n) = \sum_{n=0}^{\infty} a_{np} q^n.$$

However, this operator is not (in general) stable in the space $M_k(N,B,\rho)$, but we still have the inclusion $pU_p(M_k(N,B,\rho)) \subset M_k(N,B,\rho^p)$. To illustrate this, consider the special case of weight $k=0$. In this case, forgetting about the differential, the level structure and the growth condition in the source of f for illustrative purposes, the operator U_p can be simply written as

$$U_p(f)(E) = \frac{1}{p} \times \sum_{\substack{H \leq E[p] \\ |H|=p \\ H \neq \text{ canonical subgroup}^{11}}} f(E/H).$$

This formula for the U_p operator comes from its geometric description as $\frac{1}{p}$ times the trace of the Frobenius map[12]. Now, for H not canonical, it can be shown that if f is ρ-overconvergent for ρ such that $v_p(\rho) = r \leq \frac{1}{p+1}$, then $v_p(E_{p-1}(E/H)) = \frac{1}{p} v_p(E_{p-1}(E))$, so that $pU_p(f)$ is defined on elliptic curves $\widetilde{E}$ with $v_p(E_{p-1}(\widetilde{E})) \leq pr = v_p(\rho^p)$ and therefore $pU_p(f)$ is ρ^p-overconvergent.

Even though we fixed the weight $k=0$, the result $pU_p(M_k(N,B,\rho)) \subset M_k(N,B,\rho^p)$ is true for every weight k, provided we take $v_p(\rho) \leq \frac{1}{p+1}$ (the proof for arbitrary k can be seen in [11, III, Lemma 3.11.4]). Therefore, for the remainder of this project, we will always work with $\rho \in B$ such that $v_p(\rho) \leq \frac{1}{p+1}$. Now, since we have a natural inclusion of spaces $M_k(N,B,\rho_1) \overset{\iota}{\hookrightarrow} M_k(N,B,\rho_2)$ if ρ_2 divides ρ_1 (this is immediate from the Katz expansion of an overconvergent modular form), we can compose

$$M_k(N,B,\rho) \xrightarrow{pU_p} M_k(N,B,\rho^p) \overset{\iota}{\hookrightarrow} M_k(N,B,\rho)$$

to get a well-defined operator on $M_k(N,B,\rho)$.

Definition 3.13. *We define the p-adic Banach space $M_k(N,K,\rho)$ over K by*

[11] The canonical subgroup is a very naturally determined subgroup of $E[p] \simeq \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. For a detailed exposition about this subgroup, see Emerton's appendix in [5].

[12] This is the linear map defined on q-expansions of overconvergent modular forms by $q \mapsto q^p$. The interested reader can learn more about this viewpoint of U_p as the trace of the Frobenius in [11, III, section 3.11] or [8, pg. 43].

$$M_k(N,K,\rho) = M_k(N,B,\rho) \otimes K$$

where the norm is defined by considering $M_k(N,B,\rho)$ to be the unit ball in $M_k(N,K,\rho)$.

In order to study the eigenvalues of the U_p operator, we need a well-suited Banach space and this is the most natural choice. Notice also that this space admits a countable basis as a vector space over K, since we have a unique expression $f = \sum_{j=0}^{\infty} \rho^j \frac{b_j}{E_{p-1}^j}$ (Katz's expansion of f), where $b_j \in W_j(N,B)$. Therefore, for each $j \geq 0$ we can take a (finite) basis of the free B-module $W_j(N,B)$, divide these elements by E_{p-1}^j and join all of these sets in order to get a (countable) basis of $M_k(N,K,\rho)$ over K.

So we now have a linear map given by the composition

$$M_k(N,K,\rho) \xrightarrow{U_p} M_k(N,K,\rho^p) \xhookrightarrow{\iota} M_k(N,K,\rho)$$

which we also call U_p. A natural question is: does this linear map (on an infinite dimensional vector space) admit a trace? Of course that the answer is negative in general, but in this case we are lucky, as the following shows:

Definition 3.14. *Given two p-adic Banach spaces X, Y, let F be the subspace of continuous linear maps from X to Y whose image is finite dimensional. We say that $T : X \to Y$ is compact if T is in the closure (in the operator norm) of F.*

Theorem 3.15. *$U_p : M_k(N,K,\rho) \to M_k(N,K,\rho)$ is compact*

Proof. See Gouvêa [8, Proposition II.3.15]. □

Compact operators are very useful in p-adic functional analysis since they have well-defined traces, and even better, they support a spectral theory[13]. In particular, we can consider the Fredholm determinant $\det(1 - tU_p)$. This is a p-adic entire function[14], called the *characteristic series* of U_p, which we will denote by $P_k(t)$ (in fact, it can be shown that it does not depend on ρ, although we will not need this result). By Corollay 2.6, we can factor this series

$$P_k(t) = \textstyle\prod_{\alpha \in \mathbb{Q}} P_k(t)^{(\alpha)},$$

where $P_k(t)^{(\alpha)}$ denotes the factor of $P_k(t)$ corresponding to the characteristic roots (reciprocal of the eigenvalues) ρ_α of $P_k(t)$ with p-adic valuation $v_p(\rho_\alpha) = -\alpha$. We will also

[13] For a nice overview of the spectral theory of p-adic compact operators, see [16].

[14] This will be proved later, when we analyse the Newton polygon of this series.

denote the degree of $P_k(t)^{(\alpha)}$ by $d(k,\alpha)$ (notice the analogy with the classical case in the notation).

Let us recap what we have done so far: we have constructed a space of overconvergent modular forms with an action of the U_p operator. Moreover, this space contains the space of classical modular forms $M_k(N,B)$, by Corollary 3.12. In fact, using the canonical subgroup, one can show that the larger space $M_k(Np,B)$ of classical modular forms of weight k and level $\Gamma_1(N)\cap\Gamma_0(p)$ also embeds into $M_k(N,K,\rho)$ (for a proof of this, see [11, Theorem 3.2]). Therefore, we can conclude that

$$\mathbf{d}(k,\alpha)\leq d(k,\alpha),$$

since the left-hand side of the inequality counts generalized classical eigenforms of level $\Gamma_1(N)\cap\Gamma_0(p)$ and the right-hand side counts generalized overconvergent eigenforms of level $\Gamma_1(N)$ (both with respect to eigenvalues with p-adic valuation α). However, in order to perfectly connect the overconvergent space with the classical space, we would like to get an equality. In fact, if we require the slopes we are considering to be relatively small, we get an equality:

Theorem 3.16 (Coleman)**.** *Let $f\in M_k(N,K,\rho)$ be a generalized U_p-eigenvector with eigenvalue λ (i.e., f is in the kernel of $(U_p-\lambda)^n$ for some positive integer n) and weight k. If*

$$v_p(\lambda)<k-1,$$

then f is in $M_k(Np,B)$.

Proof. We will omit the proof, but the interested reader can see a proof of this result in [3] or in [10]. Coleman's proof uses cohomological methods applied to the space of overconvergent modular forms. Kassaei gives a more intrinsic proof of this result (and even generalizes it!) based on a gluing lemma for sections of line bundles on a rigid analytic variety. □

Corollary 3.17. *If $k>\alpha+1$, then $\mathbf{d}(k,\alpha)=d(k,\alpha)$.*

Proof. We already have $\mathbf{d}(k,\alpha)\leq d(k,\alpha)$. Since a generalized overconvergent eigenform f counted on the right-hand side is associated with an eigenvalue with p-adic valuation $\alpha<k-1$, using the previous theorem we conclude that f is classical. Hence it is also counted on the left-hand side, so $\mathbf{d}(k,\alpha)=d(k,\alpha)$. □

Therefore, in order to understand the classical quantity $\mathbf{d}(k,\alpha)$ and how it varies p-adically with the weight k, it is logical to focus our attention on the p-adic analog $d(k,\alpha)$. This is a very nice approach to the Gouvêa-Mazur conjecture and related problems since it allows us to study slopes of classical modular forms using powerful tools from p-adic functional analysis.

4 Wan's quadratic bound

Recall that for two integers k_1, k_2 such that both k_1 and k_2 are at least $2\alpha + 2$, Coleman's Theorem 3.5 tells us that there exists a finite number $\mathbf{m}(\alpha)$ such that if the congruence

$$k_1 \equiv k_2 \pmod{p^{\mathbf{m}(\alpha)}(p-1)}$$

is satisfied, then $\mathbf{d}(k_1, \alpha) = \mathbf{d}(k_2, \alpha)$. Our goal in this section is to prove Wan's quadratic bound for the quantity $\mathbf{m}(\alpha)$. Namely, we wish to prove the following theorem:

Theorem 4.1 (Wan, 1997)**.** *There are three constants A, B and C depending only on N and p such that*

$$\mathbf{m}(\alpha) \leq A\alpha^2 + B\alpha + C.$$

In order to establish this result, we will need three technical lemmas of p-adic analysis. Namely, we will first prove a continuity result stating that the two characteristic series $P_{k_1}(t)$ and $P_{k_2}(t)$ are p-adically close if k_1 and k_2 are p-adically close. Next, we will establish a uniform lower bound for the Newton polygon of $P_k(t)$ and finally we will prove a reciprocity lemma which shows that the Newton polygons of two series coincide for all the sides with small slope if the two series are p-adically close to one another.

4.1 Continuity of the characteristic series

Our goal in this section is to prove the following theorem:

Theorem 4.2. *Suppose k_1 and k_2 are integers such that $k_1 \equiv k_2 \pmod{p^n(p-1)}$ and denote $P_{k_i}(t) = \sum_{n=0}^{\infty} a_n(k_i)t^n$ the characteristic series $\det(1 - tU_p)$ of U_p acting on $M_{k_i}(N, K, \rho)$, for $i = 1, 2$. Then, for all $n \geq 0$, we have*

$$a_n(k_1) \equiv a_n(k_2) \pmod{p^{n+1}}.$$

Remark 4.3. To establish this result, we first need some observations: recall that we have a compact operator $U_p : M_k(N, K, \rho) \to M_k(N, K, \rho)$. Therefore, we can define the exterior power of this operator $\bigwedge^n U_p : \bigwedge^n M_k(N, K, \rho) \to \bigwedge^n M_k(N, K, \rho)$ for each $n \geq 0$, and, using the theory of compact operators, it can be shown that $\bigwedge^n U_p$ is also compact (see [1, Lemma 4.1]). Moreover, if $P_k(t) = \sum_{n=0}^{\infty} a_n(k)t^n$ denotes the characteristic series $\det(1 - tU_p)$, then we have $a_n(k) = (-1)^n \operatorname{Tr}(\bigwedge^n U_p)$ (see [16]).

We will also need a notion of topological $\mathbb{Z}_p$-lattice, as follows:

Definition 4.4. *Let V be a vector space over the p-adic field $\mathbb{Q}_p$. A $\mathbb{Z}_p$-submodule D of V is said to be a topological $\mathbb{Z}_p$-lattice if $V = D \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ and D is p-adically separated, i.e., we have $\bigcap_{n\geq 0} p^n D = \{0\}$.*

During this section, we will always denote the vector space $M_{k_1}(N,K,\rho)$ by V and the subspace $\{f \in M_{k_1}(N,K,\rho) | f(q) \in B[[q]]\}$ by D. In order to establish Theorem 4.2, we first need the following two lemmas:

Lemma 4.5.

1. *The q-expansion of an arbitrary element $f \in V$ has bounded denominators.*
2. *D is a topological $\mathbb{Z}_p$-lattice in V.*

Proof.

1. Write $f = \sum_{i\geq 0} \rho^i \frac{b_i}{E_{p-1}^i}$ the Katz expansion of f. Since $b_i \in B[[q]]$ tend p-adically to 0 as $i \to \infty$, there must exist some $R > 0$ such that the terms $\rho^i b_i$ are in $p^{-R}D$ for all $i \geq 0$. Now, since the q-expansion of E_{p-1}^{-i} is in D (for a proof of this fact see [8, I.2]), we conclude that f is in $p^{-R}D$ and so its q-expansion has bounded denominators.
2. It is obvious that D is a p-adically separated $\mathbb{Z}_p$-submodule of V and the tensor product condition follows immediately from the previous result, since for any $f \in V$ there exists $R > 0$ such that $p^R f \in D$.

□

Lemma 4.6. *Let Φ_1, Φ_2 be compact operators acting on the Banach space V and denote the characteristic series of the operators Φ_1 and Φ_2 by $P_1(t) = \sum_{i\geq 0} a_i t^i$ and $P_2(t) = \sum_{i\geq 0} b_i t^i$, respectively. Suppose also that the following three conditions hold:*

- *$\Phi_1(D) \subset D$,*
- *$\Phi_2(D) \subset D$,*
- *$(\Phi_1 - \Phi_2)(D) \subset p^n D$ for some non-negative integer n.*

Then, for all $i \geq 0$, we have

$$a_i \equiv b_i \pmod{p^n}.$$

Proof. We will prove the desired result by induction on $i \geq 0$.

Case i=0 Notice that by the definition of the characteristic series $P_j(t) = \det(1 - t\Phi_j)$ of Φ_j $(j = 1,2)$, it follows immediately that $a_0 = b_0 = 1$ and so obviously $a_0 \equiv b_0 \pmod{p^n}$.

Case i=1 Consider the linear operator $\Phi = \Phi_1 - \Phi_2$. Obviously Φ is compact (being the difference of two compact operators) and, by hypothesis, $\Phi(D) \subset p^n D$. Since $a_1 = -\operatorname{Tr}(\Phi_1)$ and $b_1 = -\operatorname{Tr}(\Phi_2)$, in order to conclude that $a_1 \equiv b_1 \pmod{p^n}$, we are reduced to prove that $\operatorname{Tr}(\Phi) = \operatorname{Tr}(\Phi_1) - \operatorname{Tr}(\Phi_2) \equiv 0 \pmod{p^n}$. Now, $\operatorname{Tr}(\Phi)$ is the sum of the eigenvalues λ of Φ so it is enough to prove that such an eigenvalue is $\equiv 0 \pmod{p^n}$. Let $f \in V$ be an eigenvector associated to λ and write $f = \sum\limits_{i \geq 0} c_i q^i$ with $c_i \in K$. Since the q-expansion of f has bounded denominators and K is the field of fractions of B, we can assume (replacing f by a suitable multiple if necessary) that $f \in B[[q]]$. Now, since $(\Phi_1 - \Phi_2)(D) \subset p^n D$, we have

$$\lambda \sum_{i \geq 0} a_i q^i = \lambda f = (\Phi_1 - \Phi_2)(f) = p^n \sum_{i \geq 0} b_i q^i \tag{2}$$

for some $\sum_{i \geq 0} b_i q^i \in B[[q]]$. This implies that

$$\lambda a_i = p^n b_i \tag{3}$$

for all $i \geq 0$. If there exists some a_i not divisible by p, then it follows immediately that p^n divides λ, as desired. So assume that p divides a_i for all $i \geq 0$. Thus, we must have $f \in pD$. Write $f_1 = \frac{f}{p} \in D$. Since $f_1 \in B[[q]]$ is also a Φ-eigenvector with eigenvalue λ, we can repeat the same argument in order to conclude that either p^n divides λ or $f_1 \in pD$, i.e., $f \in p^2 D$. But since D is p-adically separated, the intersection $\bigcap\limits_{n \geq 0} p^n D$ is the zero subspace $\{0\}$. Therefore, this procedure must eventually stop and so we conclude that p^n divides λ, as desired.

Inductive step Suppose that the result is true for every non-negative integer less than i. Let us now consider the linear operator $\Psi_i = \bigwedge^i \Phi_1 - \bigwedge^i \Phi_2$ on the vector space $\bigwedge^i V$, which contains the $\mathbb{Z}_p$-submodule $D_i = \bigwedge^i D$. This is in fact a topological $\mathbb{Z}_p$-lattice, since $(\bigwedge^i D) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p = \bigwedge^i (D \otimes_{\mathbb{Z}_p} \mathbb{Q}_p) = \bigwedge^i V$ and is clearly p-adically separated (since D is).

Now, to prove the inductive step, we will first prove that $\Psi_i(D_i) \subset p^n D_i$ again using induction. The case $i = 1$ is true by hypothesis. For the inductive step, simply notice that

$$\Psi_i(D_i) = [\textstyle\bigwedge^i \Phi_1 - \bigwedge^i \Phi_2](\bigwedge^i D) =$$
$$[(\textstyle\bigwedge^{i-1} \Phi_1)(\bigwedge^{i-1} D) \wedge (\Phi_1 - \Phi_2)(D)] + [(\bigwedge^{i-1} \Phi_1 - \bigwedge^{i-1} \Phi_2)(\bigwedge^{i-1} D) \wedge \Phi_2(D)]$$

which is easily seen to be the sum of two elements contained in $p^n D_i$, as $(\Phi_1 - \Phi_2)(D) \subset p^n D$ and $(\bigwedge^{i-1} \Phi_1 - \bigwedge^{i-1} \Phi_2)(D_{i-1}) \subset p^n D_{i-1}$, completing the induction argument.

Therefore, we infer that $\Psi_i(D_i) \subset p^n D_i$ and, by the same argument presented in the case $i = 1$, we conclude that $0 \equiv \mathrm{Tr}(\Psi_i) = \mathrm{Tr}(\bigwedge^i \Phi_1) - \mathrm{Tr}(\bigwedge^i \Phi_2) = (-1)^i(a_i - b_i) \pmod{p^n}$, where the last equality comes from Remark 4.3 preceding the lemma. This proves that $a_i \equiv b_i \pmod{p^n}$, as desired.

□

We will now assume k_1, k_2 to be two integers such that $k_2 - k_1 = (p-1)p^n t$ for some $t \in \mathbb{Z}$ (as in the hypotheses of Theorem 4.2), and we will denote by $\epsilon : M_{k_1}(N, K, \rho) \to M_{k_2}(N, K, \rho)$ the map that multiplies an overconvergent modular form by $E_{p-1}^{\frac{k_2-k_1}{p-1}} = E_{p-1}^{p^n t}$. Using the characterization of overconvergent modular forms via the Katz's expansion, it is easy to verify that this function is an isomorphism. Therefore, we can compose

$$M_{k_1}(N, K, \rho) \xrightarrow{\epsilon} M_{k_2}(N, K, \rho) \xrightarrow{U_p} M_{k_2}(N, K, \rho) \xrightarrow{\epsilon^{-1}} M_{k_1}(N, K, \rho)$$

to get an operator $\Phi_2 = \epsilon^{-1} U_p \epsilon$ on $M_{k_1}(N, K, \rho)$, which is also compact since the U_p operator acting on $M_{k_2}(N, K, \rho)$ is compact. Hence, as conjugate operators on a Banach space have the same characteristic series, we conclude that

$$P_{k_2}(t) = \det(1 - tU_p | M_{k_2}(N, K, \rho)) = \det(1 - t\Phi_2).$$

We also have the operator $\Phi_1 = U_p | M_{k_1}(N, K, \rho)$ and the associated characteristic series $P_{k_1}(t) = \det(1 - tU_p | M_{k_1}(N, K, \rho)) = \det(1 - t\Phi_1)$.

We can now use the previous lemma in order to establish Theorem 4.2. It is clear that the operators Φ_1 and Φ_2 defined in the previous paragraph preserve D. The only hypothesis that we still need to satisfy is $(\Phi_1 - \Phi_2)(D) \subset p^n D$. In fact, we will prove that $(\Phi_1 - \Phi_2)(D) \subset p^{n+1} D$. Noticing the decomposition

$$\Phi_2 - \Phi_1 = \epsilon^{-1} U_p \epsilon - U_p = \epsilon^{-1} U_p (\epsilon - Id_V) + (\epsilon^{-1} - Id_V) U_p, \tag{4}$$

in order to establish that $(\Phi_2 - \Phi_1)(D) \subset p^{n+1} D$ is enough to prove the following:

Lemma 4.7. *In the notation above, we have:*

1. $(\epsilon^{-1} U_p (\epsilon - Id_V))(B[[q]]) \subset p^{n+1} B[[q]]$
2. $((\epsilon^{-1} - Id_V) U_p)(B[[q]]) \subset p^{n+1} B[[q]]$

Proof.

1. Start by noticing that, by induction[15], $E_{p-1}^{p^n t} \in 1 + p^{n+1} qB[[q]]$ for all $n \geq 0$, so we can write $E_{p-1}^{p^n t} = 1 + p^{n+1} G_n$ for all $n \geq 0$ and some $G_n \in qB[[q]]$.

   Now, for $f \in B[[q]]$, we have $(\epsilon - Id_V)(f) = E_{p-1}^{\frac{k_2-k_1}{p-1}} f - f = E_{p-1}^{p^n t} f - f = (1 + p^{n+1} G_n) f - f = p^{n+1} G_n f \in p^{n+1} B[[q]]$. Since ϵ^{-1} and U_p also preserve $B[[q]]$, we conclude that $\epsilon^{-1} U_p (\epsilon - Id_V)(f) \in p^{n+1} B[[q]]$, as desired.

2. The proof of this case is completely analogous to the one above, noticing that ϵ^{-1} is multiplication by the modular form $E_{p-1}^{-p^n t}$, which is also in $1 + p^{n+1} qB[[q]]$.

□

So we know that the operator $\Phi_1 - \Phi_2$ sends $B[[q]]$ into $p^{n+1} B[[q]]$. Since it also preserves the space $M_{k_1}(N, K, \rho)$, we conclude that $(\Phi_1 - \Phi_2)(D) \subset p^{n+1} D$. Hence, we now have all of the conditions in Lemma 4.6 satisfied, so this Lemma tells us that $P_{k_1}(t) \equiv P_{k_2}(t) \pmod{p^{n+1}}$, finishing the proof of Theorem 4.2.

4.2 Lower bound for Newton polygons

In this section, we will fix $k \in \mathbb{Z}$ such that $0 \leq k < p - 1$ and establish a lower bound for the Newton polygons of the characteristic series $P_{k+j(p-1)}(t) = \det(1 - tU_p | M_{k+j(p-1)}(N, K, \rho))$ as j varies. This will be achieved by establishing first a lower bound for the p-adic valuation of the coefficients in this series.

We start by introducing some notation: Let j be a non-negative integer and denote the rank of the free B-module $M_{k+j(p-1)}(N, B)$ by d_j and the rank of the free B-module $W_j(N, B)$ (introduced in Theorem 3.9) by m_j. By Theorem 3.9, we clearly have $d_j = d_{j-1} + m_j$ for $j \geq 1$ and $d_0 = m_0$. We also write $P_{k+j(p-1)}(t) = \sum_{n=0}^{\infty} a_n (k + j(p-1)) t^n$. In order to establish the lower bound for the coefficients in this series, we will need the following three auxiliary lemmas:

Lemma 4.8. *Let V be a p-adic Banach space with an orthonormal[16] countable basis $B = \{v_i : i \geq 1\}$ and let A denote the matrix of some compact linear operator on V with respect to B. Then, we have*

$$\mathrm{Tr}(\textstyle\bigwedge^n A) = \sum_{\substack{j_1 < j_2 < \cdots < j_n \\ j_1, j_2, \cdots, j_n \in \mathbb{Z}_{\geq 1}}} [A](j_1, j_2, \cdots, j_n)$$

[15] The case $n = 0$ is well-known (the reader can see a proof in [12, Corollary 6.6]) and the inductive step is trivial.

[16] Here, orthonormal basis is meant in the following sense: every element $f \in V$ can be written in the form $f = \sum_{i \geq 1} c_i v_i$ and $\|f\| = \sup_i \{\|c_i\|\}$.

where $[A](j_1, j_2, \cdots, j_n)$ *denotes the n-th minor of A corresponding to the n rows and n lines* $j_1 < j_2 < \cdots < j_n$.

Proof. This follows from a straightforward computation using the definition of $\bigwedge^n A$ and the basis $\bigwedge^n B = \{v_{i_1} \wedge \cdots \wedge v_{i_n}\}$ of $\bigwedge^n V$. □

Lemma 4.9. *Let* V_p *be the linear map acting on q-expansions of overconvergent modular forms via* $q \mapsto q^p$ *(the Frobenius map). Then, for every* $f, g \in M_k(N, K, \rho)$*, we have*

$$U_p(g.V_p(f)) = f.U_p(g)$$

(this property is commonly called the Frobenius linearity of the operator U_p*).*

Proof. This easily follows using the action of the operators U_p and V_p on the q-expansions of f and g. □

Lemma 4.10. *Let* V_p *be the Frobenius operator and G be the overconvergent modular form* $\frac{E_{p-1}}{V_p(E_{p-1})}$*. Then G is a* 1*-unit in the ring*[17] $M_0(N, B, \rho)$ *for every* $\rho \in B$ *satisfying* $0 \leq v_p(\rho) \leq \frac{1}{p+1}$.

Proof. See Wan [17, Lemma 2.1]. □

We can now present the lower bound on the p-adic valuation of the coefficients in the characteristic series:

Proposition 4.11. *Let* $n \geq d_0$ *be an integer and suppose that* $d_l \leq n < d_{l+1}$ *for some*[18] $l \geq 0$*. Then, we have*

$$v_p(a_n(k + j(p-1))) \geq \frac{p-1}{p+1}[\sum_{u=0}^{l} u m_u + (l+1)(n - d_l)] - n$$

Proof. We start by applying a trick already used in the previous section. Namely, we conjugate the map U_p acting on $M_{k+j(p-1)}(N, B)$ by the multiplication by E_{p-1}^j map (an isomorphism of Banach spaces, as previously explained):

$$M_k(N, K, \rho) \xrightarrow{E_{p-1}^j} M_{k+j(p-1)}(N, K, \rho) \xrightarrow{U_p} M_{k+j(p-1)}(N, K, \rho) \xrightarrow{E_{p-1}^{-j}} M_k(N, K, \rho)$$

In this way, we get a map $E_{p-1}^{-j} \circ U_p \circ E_{p-1}^j : M_k(N, K, \rho) \to M_k(N, K, \rho)$ acting on a space independent of j. Now, since conjugate operators have the same characteristic series, we conclude that

[17] Notice that the space $M_0(N, B, \rho)$ is in fact a ring since multiplying two overconvergent modular forms of weight 0 gives again an overconvergent modular form of weight 0.

[18] Such an l must always exist since the quantity d_l grows with l, as we will see in Theorem 4.12.

$$P_{k+j(p-1)}(t) = \det(1 - tU_p | M_{k+j(p-1)}(N,K,\rho)) = \det(1 - tE_{p-1}^{-j} U_p E_{p-1}^{j} | M_k(N,K,\rho)).$$

Let us look more closely at the map $E_{p-1}^{-j} \circ U_p$. By Lemma 4.9, we know $E_{p-1}^{-j} \circ U_p = U_p(V_p(E_{p-1}^{-j})) = U_p \circ E_{p-1}^{-j}(q^p)$ and so we have $E_{p-1}^{-j} \circ U_p \circ E_{p-1}^{j} = U_p(E_{p-1}^{-j}(q^p)) \circ E_{p-1}^{j} = U_p \circ (\frac{E_{p-1}}{V_p(E_{p-1})})^j$. Therefore, we have

$$P_{k+j(p-1)}(t) = \det(1 - tE_{p-1}^{-j} U_p E_{p-1}^{j} | M_k(N,K,\rho)) = \det(1 - tU_p \circ (\frac{E_{p-1}}{V_p(E_{p-1})})^j) = \det(1 - tU_p \circ G^j)$$

where G is the map given by multiplication with the overconvergent modular form $\frac{E_{p-1}}{V_p(E_{p-1})}$ of Lemma 4.10. Hence, we are reduced to the study of the characteristic series of the compact[19] operator $U_p \circ G^j$ in the space $M_k(N,K,\rho)$ (which does not depend on j). In order to study it properly, we introduce some notation: for every $i \geq 0$, let $\{b_{i,1}, \cdots, b_{i,m_i}\}$ be a basis of the free B-module $W_i(N,B)$. Now, using the Katz expansion $f = \sum_{j=0}^{\infty} \rho^j \frac{b_j}{E_{p-1}^j}$ of an arbitrary element $f \in M_k(N,K,\rho)$, we conclude that the elements

$$e_{i,s} = \rho^i \frac{b_{i,s}}{E_{p-1}^i}$$

form an orthonormal basis of $M_k(N,K,\rho)$. Therefore, for each $i \geq 0$ and $1 \leq s \leq m_i$, we can write

$$U_p \circ G^j(e_{i,s}) = \sum_{u,v} A_{i,s}^{u,v}(j) e_{u,v}$$

with $A_{i,s}^{u,v}(j) \in K$. Now, notice that the space $M_k(N,B,\rho)$ is a module over $M_0(N,B,\rho)$ and Lemma 4.10 tells us that G^j is a unit in $M_0(N,B,\rho)$. Therefore, we conclude that multiplication by G^j is stable on $M_k(N,B,\rho)$ (the unit ball of $M_k(N,K,\rho)$) and so the elements $G^j(e_{i,s})$ are in $M_k(N,B,\rho)$. Using the result $pU_p(M_k(N,B,\rho)) \subset M_k(N,B,\rho^p)$ of section 3.3, we infer that

$$U_p \circ G^j(e_{i,s}) = \frac{1}{p} \sum_u \frac{(\rho^p)^u}{E_{p-1}^u} b_u(i,s,j) = \frac{1}{p} \sum_u \rho^{(p-1)u} (\frac{\rho}{E_{p-1}})^u b_u(i,s,j)$$

for some classical modular forms $b_u(i,s,j) \in W_u(N,B)$. Writing $b_u(i,s,j) = \sum_{t=1}^{m_u} c_{t,u,i,s,j} b_{u,t}$ for some $c_{t,u,i,s,j} \in B$, we can also express $U_p \circ G^j(e_{i,s})$ as

$$\frac{1}{p} \sum_{u,t} \rho^{(p-1)u} (\frac{\rho}{E_{p-1}})^u c_{t,u,i,s,j} b_{u,t} = \frac{1}{p} \sum_{u,t} \rho^{(p-1)u} c_{t,u,i,s,j} e_{u,t}.$$

[19] This operator is compact since U_p is compact. For more details on this, see [16, pg. 72].

Therefore, comparing this last expression with $U_p \circ G^j(e_{i,s}) = \sum_{u,v} A_{i,s}^{u,v}(j) e_{u,v}$, we conclude that $A_{i,s}^{u,v}(j) = \frac{\rho^{(p-1)u}}{p} c_{v,u,i,s,j}$, so we immediately get the bound

$$v_p(A_{i,s}^{u,v}(j)) \geq u(p-1)v_p(\rho) - 1 \tag{5}$$

This is a lower bound for the coefficients of the (infinite) matrix of $U_p \circ G^j$. Our goal now is to convert it into a lower bound for the coefficients of the corresponding characteristic series. As it was explained in the previous section, we have $a_n(k + j(p-1)) = (-1)^{k+j(p-1)} \operatorname{Tr}(\bigwedge^n (U_p \circ G^j))$. Hence, we are in perfect conditions to apply Lemma 4.8. In fact, if we consider the basis $\{e_{i,s}\}$ of $M_k(N, K, \rho)$ and denote by $A = (a_{ij})$ the infinite matrix of $U_p \circ G^j$ with respect to this basis, Lemma 4.8 tells us that $\operatorname{Tr}(\bigwedge^n (U_p \circ G^j)) = \sum_{\substack{j_1 < j_2 < \cdots < j_n \\ j_1, j_2, \cdots, j_n \in \mathbb{Z}_{\geq 1}}} [A](j_1, j_2, \cdots, j_n)$. In particular, we have

$$\begin{aligned} v_p(a_n(k + j(p-1))) &= v_p\Big(\sum_{\substack{j_1 < j_2 < \cdots < j_n \\ j_1, j_2, \cdots, j_n \in \mathbb{Z}_{\geq 1}}} [A](j_1, j_2, \cdots, j_n) \Big) \\ &\geq \inf_{\substack{j_1 < j_2 < \cdots < j_n \\ j_1, j_2, \cdots, j_n \in \mathbb{Z}_{\geq 1}}} v_p([A](j_1, j_2, \cdots, j_n)). \end{aligned}$$

Also, looking at the bound (5), we see that the p-adic valuation of the coefficients of the matrix A are bounded constantly in the lines corresponding to basis vectors $e_{u,v}$ with a fixed u and varying $1 \leq v \leq m_u$, and grow (linearly) with u. Thus, we have the following division of A

$$A = \begin{bmatrix} A_0 \\ A_1 \\ \cdots \\ A_u \\ \cdots \end{bmatrix}$$

where A_u is the (infinite) submatrix of A consisting of the m_u lines that correspond to the vectors $e_{u,1}, \cdots, e_{u,m_u}$. In each one of these submatrices the bound (5) is constant and therefore it is easy to see that the desired infimum is attained in some square submatrix consisting of the first n lines of A.

Now, in order to make the calculation effective, notice that, for $j \geq 1$, the relation $d_j = d_{j-1} + m_j$ implies $d_j = \sum_{i=0}^{j} m_i$, so that if we locate n in the interval of integers $[d_l, d_{l+1}]$, then the first n lines will consist of $m_0 + m_1 + \cdots + m_l + (n - d_l)$ lines (where the final $(n - d_l)$ lines are in fact the first $(n - d_l)$ lines of the submatrix A_{l+1}). Therefore, using the bound (5) for the coefficients of each one of the submatrices A_u, we get

$$v_p(a_n(k+j(p-1))) \geq v_p([A](1,2,\cdots,n)) \geq$$
$$[\sum_{u=0}^{l} u(p-1)v_p(\rho)m_u - m_u] + (l+1)(p-1)v_p(\rho)(n-d_l) - (n-d_l) =$$
$$v_p(\rho)(p-1)[\sum_{u=0}^{l} um_u + (l+1)(n-d_l)] - n$$

where in the second inequality we used the permutation expansion formula for the determinant and in the last line we used the identity $d_l = \sum_{i=0}^{l} m_i$. Since this is true for every $\rho \in B$ such that $0 \leq v_p(\rho) \leq \frac{1}{p+1}$, we conclude that $v_p(a_n(k+j(p-1))) \geq \frac{p-1}{p+1}[\sum_{u=0}^{l} um_u + (l+1)(n-d_l)] - n$, which is the inequality we were trying to prove.

□

Our next goal is to transform this bound into a quadratic bound on n. In order to do this, we will need to control how the quantity m_j in the previous bound varies with j. In fact, we will use the following dimension formulas for the rank of the free B-module $M_{k+j(p-1)}(N,B)$ of classical modular forms:

Theorem 4.12 (Dimension formulas)**.** *Let k be a fixed non-negative integer, $\Gamma = \Gamma_1(N)$, g the genus of $X(\Gamma)$, ϵ_2 the number of elliptic points[20] with period 2 and ϵ_3 the number of elliptic points with period 3. Also, let ϵ_∞ be the number of cusps, ϵ_∞^{reg} the number of regular cusps and ϵ_∞^{irr} the number of irregular cusps. Then, the rank of the B-module $M_{k+j(p-1)}(N,B)$ of classical modular forms of weight $w := k+j(p-1)$ and level $\Gamma_1(N)$ is given by:*

$$d_j = \begin{cases} (w-1)(g-1) + \lfloor \frac{w}{4} \rfloor \epsilon_2 + \lfloor \frac{w}{3} \rfloor \epsilon_3 + \frac{w}{2}\epsilon_\infty & \text{if } w \geq 2 \text{ is even} \\ (w-1)(g-1) + \lfloor \frac{w}{3} \rfloor \epsilon_3 + \frac{w}{2}\epsilon_\infty^{reg} + \frac{w-1}{2}\epsilon_\infty^{irr} & \text{if } w \geq 3 \text{ is odd and } -I \text{ is not}^{21} \text{in } \Gamma \end{cases}$$

Proof. See Diamond and Shurman [6, III, Theorems 3.5.1 and 3.6.1]. □

Corollary 4.13. *There exists a constant α (only depending on N, p and k) such that*

$$m_j \geq \alpha$$

for all $j \geq 0$.

Proof. Using the previous theorem, it is clear that we can find three constants A, B, C (with $A > 0$) that only depend on the congruence subgroup $\Gamma_1(N)$ (and hence, depending only on N), such that

[20] The definition of elliptic point and regular/irregular cusp can be seen in [6, III].

[21] If the congruence subgroup Γ contains the negative identity matrix $-I$, then clearly $M_w(N,B) = \{0\}$.

$$A(k+j(p-1))+B \geq d_j \geq A(k+j(p-1))+C.$$

for all $j \geq 1$. Now, recall that $m_j = d_j - d_{j-1}$ for $j \geq 1$. Hence, using this last inequality, we get

$$m_j = d_j - d_{j-1} \geq A(p-1)+C-B.$$

The result follows putting $\alpha = \max\{m_0, A(p-1)+C-B\}$.

□

Corollary 4.14. *There exists three constants a,b,c (only depending on N,p and k) with $a>0$ such that*

$$v_p(a_n(k+j(p-1))) \geq an^2+bn+c$$

for all $n \geq 1$.

Proof. By Corollary 4.13 there exists a constant α such that $m_u \geq \alpha$ and the dimension formulas also give us three constants A,B,C such that

$$A(k+l(p-1))+B \geq d_l \geq A(k+l(p-1))+C.$$

In particular, if $n \geq d_0$, we can locate $d_l \leq n < d_{l+1}$, and so clearly both numbers l and $n-d_l$ are bounded (above and below) by some linear function in n, whose coefficients only depend on N,p and k. Therefore, using the bound of Proposition 4.11, we get

$$v_p(a_n(k+j(p-1))) \geq \tfrac{p-1}{p+1}[\alpha \sum_{u=0}^{l} u + (l+1)(n-d_l)] - n \geq \tfrac{p-1}{p+1}[\alpha \tfrac{l(l+1)}{2} + (l+1)(n-d_l)] - n,$$

which is clearly a quadratic bound on n whose coefficients depend only on N,p and k, since both l and d_l are controlled by linear functions in n whose coefficients only depend on these three parameters. This inequality is valid for all $n \geq d_0$ and adjusting the coefficients in this quadratic bound if necessary, we can assume that it is also valid for all $n \geq 1$, as desired.

□

4.3 Reciprocity Lemma

Our goal in this section is to transform the quadratic bound of the last section into an upper bound for the quantity $\mathbf{m}(\alpha)$. For k_1, k_2 integers such that $k_2 = k_1 + j(p-1)$ (for some $j \in \mathbb{Z}$), we again denote the characteristic series of the U_p map acting on $M_{k_i}(N, K, \rho)$ by $P_{k_i}(t)$ for $i = 1, 2$. We now present the reciprocity lemma that will allow us to construct the desired bound.

Lemma 4.15. *Let $N_i(x)$ be the function defined on $\mathbb{R}_{\geq 0}$ whose graph is the Newton polygon[22] of $P_{k_i}(t)$ and let $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be a function satisfying the following conditions:*

(a) $f(x)$ *is continuous and strictly increasing,*

(b) $f(0) \leq 0$,

(c) $xf(x) \leq N_i(x)\ \forall x \geq 1$ *and* $i = 1, 2$,

(d) $\lim_{x \to +\infty} f(x) = +\infty$,

(e) $xf^{-1}(x)$ *is an increasing function*[23].

Define also the increasing function (on $\mathbb{R}_{\geq 0}$) $m_f(x) = \lfloor xf^{-1}(x) \rfloor$. Now, if the congruence

$$P_{k_1}(t) \equiv P_{k_2}(t) \pmod{p^{m_f(\alpha)+1}}$$

is true for some $\alpha \geq 0$, then the Newton polygons of $N_1(x)$ and $N_2(x)$ coincide for all sides with slopes at most α.

Proof. For $i = 1, 2$, let γ be a slope of some side of the Newton polygon $N_i(x)$ and suppose this side has an end-vertex $(n, N_i(n))$ for some natural number $n \geq 1$. Now, by the geometric construction of Newton polygons, it is clear that $N_i(n) \leq \gamma n$ and so we have

$$\gamma \geq \frac{N_i(n)}{n} \overset{(c)}{\geq} \frac{nf(n)}{n} = f(n). \tag{6}$$

Applying the increasing function $f^{-1}(x)$ to the inequality $\gamma \geq f(n)$ gives

$$f^{-1}(\gamma) \geq n. \tag{7}$$

Hence, using the first inequality in (6) and the previous inequality (7), we conclude that

$$N_i(n) \leq \gamma n \leq \gamma f^{-1}(\gamma). \tag{8}$$

[22] Notice that we can define these Newton polygons since $P_{k_1}(0) = P_{k_2}(0) = 1$ so that the constant term in both series is 1.

[23] Notice that this function is well-defined on $\mathbb{R}_{\geq 0}$ by the previous hypotheses.

Now, by the hypothesis (d), the function $xf(x)$ must grow faster than any linear function. Therefore, since $N_i(x) \geq xf(x)$, the slopes of the sides of the two Newton polygons $N_i(x)$ tend to ∞ as $x \to \infty$. In particular, using Proposition 2.4, we conclude that $P_{k_1}(t)$ and $P_{k_2}(t)$ are p-adic entire functions.

We now prove the desired conclusion of the Lemma: let $\alpha_1 < \cdots < \alpha_s$ be the non-negative rational numbers which are at most α and which occur as a slope of some side in at least one of the two Newton polygons $N_i(x)$ (notice that any slope of $N_i(x)$ is non-negative by hypothesis (c)). Let us prove by induction on $s \geq 1$ that the two Newton polygons $N_i(x)$ coincide for all the sides with slopes at most α_s. The case $s = 1$ is trivial. Suppose now that the two Newton polygons $N_i(x)$ coincide for all the sides with slopes at most α_{s-1} and assume, without loss of generality, that α_s appears as a slope of $N_1(x)$ with endpoint $(n_s, N_1(n_s))$. Let β be the slope of $N_2(x)$ occurring after α_{s-1} (if such a β does not exist, set $\beta = \infty$). We want to prove that $\beta = \alpha$. Clearly $\beta \geq \alpha_s$ and by inequality (8) and the fact that $m_f(x)$ is increasing, we have:

$$N_1(n_s) \overset{(8)}{\leq} \alpha_s f^{-1}(\alpha_s) < m_f(\alpha_s) + 1 \leq m_f(\alpha) + 1. \tag{9}$$

Now, if we write $P_{k_i}(t) = 1 + \sum_{n=1}^{\infty} a_n(i)t^n$, the congruence hypothesis in the Lemma assure us that

$$a_{n_s}(1) \equiv a_{n_s}(2) \pmod{p^{m_f(\alpha)+1}}. \tag{10}$$

Furthermore, since the point $(n_s, N_1(n_s))$ is in the graph of $N_1(x)$, by the definition of Newton polygon we must have:

$$v_p(a_{n_s}(1)) = N_1(n_s).$$

Therefore, since by inequality (9) we have $v_p(a_{n_s}(1)) = N_1(n_s) < m_f(\alpha) + 1$, the congruence relation (10) allows us to infer that

$$v_p(a_{n_s}(2)) = N_1(n_s).$$

This proves that $(n_s, N_1(n_s)) = (n_s, N_2(n_s))$. Since by the inductive hypothesis $N_1(x)$ and $N_2(x)$ coincide for all the sides with slope at most α_{s-1} and $(n_s, N_1(n_s))$ is a common point in a side of slope α_s, we must have $\beta \leq \alpha_s$. Hence, we have $\beta = \alpha_s$. Now, clearly the side of slope α_s of $N_2(x)$ is at least as long as the side of slope α_s of $N_1(x)$ (since $(n_s, N_1(n_s)) = (n_s, N_2(n_s))$ is the endpoint of this side in $N_1(x)$). We can now apply a symmetric argument with $N_2(x)$ (since we now know that $\beta = \alpha_s$) to conclude that in fact

this side has the same length in $N_1(x)$ and $N_2(x)$. This finishes the induction step and the proof of the lemma.

□

This lemma allows us to construct an upper bound for the quantity $\mathbf{m}(\alpha)$ in the following way: by Corollary 4.14, there are three constants a, b, c (depending on N, p and k) with $a > 0$ such that

$$N_i(x) \geq ax^2 + bx + c$$

for $x \geq 1$ and $i = 1,2$. Now, define $\widetilde{m}(x) = \lfloor \frac{x(x+|b|+|c|)}{a} \rfloor$ and $f(x) = ax - |b| - |c|$. It is easy to check that $f(x)$ satisfies all the conditions in the previous lemma and that $xf^{-1}(x) = \frac{x(x+|b|+|c|)}{a}$ (so that $\widetilde{m}(x) = m_f(x)$ in the notation of the lemma). Now, suppose that k_1, k_2 are two integers (both at least $2\alpha + 2$) such that $k_1 \equiv k_2 \pmod{p^{\widetilde{m}(\alpha)}(p-1)}$. Then Theorem 4.2 tells us that $P_{k_1}(t) \equiv P_{k_2}(t) \pmod{p^{\widetilde{m}(\alpha)+1}}$. Hence, applying the previous lemma, we conclude that the two Newton polygons $N_1(x)$ and $N_2(x)$ coincide for all the sides with slopes at most α. In particular, using the power series version of Proposition 2.2 (see [7, Corollary 6.5.11], for example), the number of roots (with multiplicity) of $P_{k_i}(t)$ with p-adic valuation α is the same for $i = 1,2$. Thus, we conclude that $\mathbf{d}(k_1, \alpha) = \mathbf{d}(k_2, \alpha)$. We are almost ready to prove Wan's Theorem. But to do so, we need a final concept:

Definition 4.16. *We define the non-negative integer $m(\alpha, k)$ to be the smallest integer*[24] *such that for all non-negative integers k_1, k_2 in the residue class k modulo $p-1$ satisfying $k_1 \equiv k_2 \pmod{p^{m(\alpha,k)}(p-1)}$, we have the equality $\mathbf{d}(k_1, \alpha) = \mathbf{d}(k_2, \alpha)$.*

Let us recap what we have proved so far: we have established that for $k_1 \equiv k_2 \pmod{p^{\widetilde{m}(\alpha)}(p-1)}$, we have $\mathbf{d}(k_1, \alpha) = \mathbf{d}(k_2, \alpha)$. But since $m(\alpha, k)$ is the smallest value that we can set in the congruence $k_1 \equiv k_2 \pmod{p^{m(\alpha,k)}(p-1)}$ in order to have $\mathbf{d}(k_1, \alpha) = \mathbf{d}(k_2, \alpha)$, we conclude that $m(\alpha, k) \leq \widetilde{m}(\alpha)$, a quadratic polynomial on α. Moreover, since clearly $\mathbf{m}(\alpha) = \max_{0 \leq k < p-1} m(\alpha, k)$, we conclude that $\mathbf{m}(\alpha)$ is dominated by a quadratic bound on α which only depends on N and p (notice that this bound will not depend on k now since we took the maximal bound for $0 \leq k < p-1$), establishing Wan's Theorem 4.1.

4.4 A concrete example

In this section we will explicitly compute the quadratic bound for $\mathbf{m}(\alpha)$ obtained in the end of the previous section for the special case $N = 1$. We will also work with the concrete prime number $p = 73$, although the argument can easily be adapted for a prime $p \equiv 1 \pmod{12}$ and even generalized for an arbitrary prime (but perhaps with weaker bounds).

[24] Coleman's Theorem 3.5 guarantees that such an integer always exists.

We will start by fixing a weight k such that $0 \leq k < 72$. Using the dimension formulas for the special case $N = 1$, we know that

$$d_l = \begin{cases} \lfloor \frac{k}{12} \rfloor + 6l + 1 & \text{if } k \not\equiv 2 \pmod{12} \\ \lfloor \frac{k}{12} \rfloor + 6l & \text{if } k \equiv 2 \pmod{12} \end{cases}$$

and looking at this formula it is also easy to conclude that the function d_l satisfies

$$d_l = d_0 + 6l \tag{11}$$

for $l \geq 0$. Now, in order to get the desired bound, suppose that n is a positive integer such that $d_l \leq n < d_{l+1}$, as in the hypotheses of Proposition 4.11. Given an integer $u \geq 1$, the dimension formula also implies that

$$m_u = d_u - d_{u-1} = 6. \tag{12}$$

Therefore, identities (11) and (12) together with Proposition 4.11 for $p = 73$ give the bound

$$\begin{aligned} v_{73}(a_n(k+72j)) &\overset{4.11}{\geq} \frac{72}{74}\Big(\sum_{u=0}^{l} u m_u + (l+1)(n-d_l)\Big) - n \\ &\overset{(12)}{=} \frac{72}{74}\Big(\frac{6l(l+1)}{2} + (l+1)(n-d_l)\Big) - n \\ &= \frac{72}{74}(l+1)\Big(\frac{6l}{2} + (n-d_l)\Big) - n \\ &\overset{(11)}{=} \frac{72}{74}(l+1)\Big(\frac{6l}{2} + (n-(d_0+6l))\Big) - n \\ &= \frac{72}{74}(l+1)(n-d_0-3l) - n \end{aligned} \tag{13}$$

Now, since we are assuming $d_l \leq n < d_{l+1}$, using the dimension formula (11) we also have the bound for l

$$\frac{n-d_0}{6} - 1 < l \leq \frac{n-d_0}{6}. \tag{14}$$

We can now use this identity in the last line of the inequality (13) to get

$$\begin{aligned} v_{73}(a_n(k+72j)) =& \frac{72}{74}(l+1)(n-d_0-3l)-n \\ \overset{(14)}{\geq}& \frac{72}{74}(\frac{n-d_0}{6})(n-d_0-3\frac{n-d_0}{6})-n \\ =& \frac{72}{74}(\frac{n-d_0}{6})(\frac{n-d_0}{2})-n \end{aligned}$$

This give us the desired quadratic bound[25] $v_{73}(a_n(k+72j)) \geq An^2+Bn+C$, where A,B,C are explicitly given by:

- $A = \frac{3}{37}$
- $B = -\frac{6}{37}d_0 - 1$
- $C = \frac{3}{37}d_0^2$

Now, recall that we began our calculation with a positive integer n located in the interval $[d_l, d_{l+1}]$. However, if we forget the positive constant term $\frac{3}{37}d_0^2$ in this last bound, we obtain the weaker (but still valid) inequality

$$v_{73}(a_n(k+72j)) \geq \tfrac{3}{37}n^2 + (-\tfrac{6}{37}d_0 - 1)n.$$

Analysing the parabola given by this quadratic bound, it is easy to see that the quadratic polynomial on the right-hand-side is non-positive for $0 \leq n \leq d_0$ and so we have a bound that is actually valid for all $n \geq 0$ and for a fixed weight in the residue class k modulo 72 (recall that we took k such that $0 \leq k < 72$).

Now it is time to use the work of the previous sections in order to transform this last quadratic bound into an upper bound for the quantity $\mathbf{m}(\alpha)$. Recall that the quantity $m(\alpha,k)$ represents the smallest integer such that for all non-negative integers k_1, k_2 in the residue class k modulo 72 satisfying $k_1 \equiv k_2 \pmod{73^{m(\alpha,k)}72}$, we have the equality $\mathbf{d}(k_1,\alpha) = \mathbf{d}(k_2,\alpha)$. Therefore, by the work done in the end of section 4.3, we infer the bound

$$m(\alpha,k) \leq \frac{37}{3}\alpha^2 + (2d_0 + \frac{37}{3})\alpha.$$

Finally, since $\mathbf{m}(\alpha) = \max_{0\leq k<72} m(\alpha,k)$, it is clear that the desired quadratic bound is attained when $d_0 = \lfloor \frac{k}{12} \rfloor + 1$ (for $k \not\equiv 2 \pmod{12}$) is maximal for $k < 72$. This obviously happens when $d_0 = 6$ and so we can give the final bound for $\mathbf{m}(\alpha)$:

$$\mathbf{m}(\alpha) \leq 12.34\alpha^2 + 24.34\alpha$$

[25] Notice that this bound only depends on d_0, the dimension of the vector space $M_k(1,B)$, and hence only on k.

References

[1] M. Bessa, On the spectrum of generic random product of compact operators, 2006.

[2] K. Buzzard and F. Calegari, A counterexample to the Gouvêa-Mazur conjecture, 2003, arXiv:math/0311361v1 [math.NT].

[3] R. Coleman, Classical and oveconvergent modular forms, Invent. Math., 124 (1996), 215-241.

[4] R. Coleman, *p*-adic Banach spaces and families of modular forms, Invent. Math., 127 (1997), 417-479.

[5] B. Conrad and K. Rubin, Arithmetic Algebraic Geometry, American Mathematical Society, IAS/Park City Mathematics Institute (2008).

[6] F. Diamond and J. Shurman, *A First Course in Modular Forms*, Springer-Verlag, Graduate Texts in Mathematics, No. 228, New York, 2005.

[7] F. Gouvêa, *p-adic Numbers*, Springer-Verlag Berlin Heidelberg, Universitext, 1997.

[8] F. Gouvêa, *Arithmetic of p-adic Modular Forms*, Springer-Verlag, Lecture Notes in Mathematics, Vol 1304 (1988).

[9] F. Gouvêa and B. Mazur, Families of modular eigenforms, Math. Comp., Vol. 58, No. 198 (1992), 793-805.

[10] P. Kassaei, A gluing lemma and overconvergent modular forms, Duke Math. J. 132 (2006), no. 3, 509–529.

[11] N. Katz, *p*-adic properties of modular schemes and modular forms, in Modular Forms in One Variable III (SLN 350), Springer-Verlag, 1973, 69-190.

[12] L. Kilford, *Modular Forms: A Classical and Computational Introduction*, Imperial College Press, 2nd edition.

[13] N. Koblitz, *p-adic numbers, p-adic Analysis, and Zeta-Functions*, Springer-Verlag, Graduate Texts in Mathematics, No. 58, New York, 1984.

[14] A. Lauder, Computations with classical and *p*-adic modular forms, LMS J. Comput. Math. 14, (2011), 214-231.

[15] J-P. Serre, Formes modulaires et fonctiones zeta *p*-adiques, in Modular Forms in One Variable III (SLN 350) 1973, Springer-Verlag, 191-268.

[16] J-P. Serre, Endomorphismes complètement continus des espaces de Banach *p*-adiques, in Publications Mathématiques de l'Institut des Hautes Études Scientifiques, December 1962, Volume 12, Issue 1, 69-85.

[17] D. Wan, Dimension variation of classical and p-adic modular forms, Invent. Math., 133 (1998), 449-463.