On the semi-simplicity of the U_p-operator on modular forms.

Robert F. Coleman   Bas Edixhoven*

Fall 1996

1 Introduction.

For N and k positive integers, let $M^0(N,k)_{\mathbb{C}}$ denote the $\mathbb{C}$-vector space of cuspidal modular forms of level N and weight k. This vector space is equipped with the usual Hecke operators T_n, $n \geq 1$. If we need to consider several levels or weights at the same time, we will denote this T_n by T_n^N, or $T_n^{N,k}$. If p is a prime number dividing N, our T_p is also known under the name U_p. One of our main results can be stated very easily: if $k = 2$ and p^3 does not divide N, then the operator T_p is semi-simple. We can prove the same result for weight $k \geq 3$, under the assumption that certain crystalline Frobenius elements are semi-simple. Milne has shown in [11, §2] that this semi-simplicity is implied by Tate's conjecture claiming that for X projective and smooth over a finite field of characteristic p, and $r \geq 0$, $\dim_{\mathbb{Q}}(\mathrm{CH}^r(X)/\equiv_{\mathrm{num}})$ equals the order of $\zeta(X,s)$ at r. Ulmer proved in [17] that T_p is semi-simple, for $k = 3$ and p^2 not dividing N, under the assumption of the Birch-Swinnerton-Dyer conjecture for elliptic curves over function fields in characteristic p. His method is quite different from ours: assuming that T_p is not semi-simple, he really shows that the Birch-Swinnerton-Dyer conjecture does not hold for an explicitly given elliptic curve.

The structure of our proof is as follows. Using the theory of newforms, the problem is shown to be equivalent to the problem of showing that, for a normalized newform f of weight k, prime-to-p level and character ε, the polynomial $x^2 - a_p x + \varepsilon(p)p^{k-1}$ has no double root. This polynomial happens to be the characteristic polynomial of the Frobenius element at p in the two-dimensional Galois representations associated to f; it is also the characteristic polynomial of the crystalline Frobenius asociated to f. We show that this crystalline Frobenius cannot be a scalar.

In Sections 2 and 3 we prove the results concerning these Frobenius elements for $k = 2$ and $k \geq 2$, respectively. Section 2 is quite elementary, whereas in Section 3 we use a lot of the machinary for comparing p-adic etale and crystalline cohomology. In Section 4 we give some applications: the Ramanujan inequality is a strict inequality in certain cases, certain Hecke algebras are reduced, hence have non-zero discriminant. Section 5 gives some results, due to Abbes and Ullmo, concerning the discriminants of certain Hecke algebras.

*partially supported by the Institut Universitaire de France

To end this introduction, let us explain why the case $k = 1$ is completely different. Consider a normalized cuspidal eigenform $f = \sum a_n q^n$ of some level N, of weight one and with some character ε. Deligne and Serre have shown ([5, §4]) that there exists a continuous representation ρ_f from $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ to $\mathrm{GL}_2(\mathbb{C})$, unramified outside N, such that, for all primes p not dividing N, the characteristic polynomial of a Frobenius element at p is $x^2 - a_p x + \varepsilon(p)$. Since the image of ρ_f is finite, Chebotarev's density theorem gives the existence of primes p not dividing N such that the Frobenius element at p is the identity element, hence has characteristic polynomial $(x-1)^2$.

2 An elementary proof in the case of weight two.

2.1 Theorem. *Let $f = \sum a_n q^n$ be a cuspidal normalized eigenform of weight two, some level N and character $\varepsilon\colon (\mathbb{Z}/N\mathbb{Z})^* \to \mathbb{C}^*$. Let p be a prime number not dividing N. Then the polynomial $x^2 - a_p x + \varepsilon(p)p$ has simple roots.*

Proof. The proof is by contradiction, so we suppose that the polynomial has a double root λ. Then of course we have $\lambda^2 = \varepsilon(p)p$ and $2\lambda = a_p$. Let K be the finite extension of $\mathbb{Q}$ generated by the a_n and the $\varepsilon(a)$, and let O_K be its ring of integers. Let J denote the jacobian variety of the modular curve $X_1(N)$ over $\mathbb{Q}$. We identify the space $M^0(N,2)_{\mathbb{Q}}$ of weight two cuspforms of level N and with coefficients in $\mathbb{Q}$ with the cotangent space at the origin of J; this is compatible with the action of the Hecke operators. Let $\mathbb{T}$ be the subring of $\mathrm{End}(J)$ that is generated by the T_n, $n \geq 1$, and the diamond operators $\langle a\rangle$, $a \in (\mathbb{Z}/N\mathbb{Z})^*$. Let I be the annihilator of f in $\mathbb{T}$, and let $A'_{\mathbb{Q}} := J/IJ$ be the quotient of J by its subvariety generated by the images of all elements of I. It is well known that $A'_{\mathbb{Q}}$ has dimension $[K:\mathbb{Q}]$, and that for every prime number l, the free $K \otimes \mathbb{Q}_l$-module $V_l(A'_{\mathbb{Q}})$ of rank two gives the l-adic Galois representation $\rho_{f,l}$ associated to f. We prefer to work with an abelian variety $A_{\mathbb{Q}}$ that is isogeneous to $A'_{\mathbb{Q}}$ and on which we have an action of all of O_K. This is easily done: define $A_{\mathbb{Q}} := O_K \otimes_{\mathbb{T}} A'_{\mathbb{Q}}$, where the tensor product should be calculated by taking a presentation of O_K.

The abelian variety $A_{\mathbb{Q}}$ has good reduction at p; let $A_{\mathbb{Z}_p}$ denote the corresponding abelian scheme over $\mathbb{Z}_p$. We consider the first algebraic de Rham cohomology group $M := \mathrm{H}^1_{\mathrm{DR}}(A_{\mathbb{Z}_p}/\mathbb{Z}_p)$. It is a free $\mathbb{Z}_p$-module of rank $2[K:\mathbb{Q}]$, equipped with its Hodge filtration:

$$M = \mathrm{Fil}^0 M \supset \mathrm{Fil}^1 M = \mathrm{H}^0(A_{\mathbb{Z}_p}, \Omega^1). \tag{2.2}$$

The submodule $\mathrm{Fil}^1 M$ is free of rank $[K:\mathbb{Q}]$ as $\mathbb{Z}_p$-module, and has the property that $\mathrm{Fil}^0 M/\mathrm{Fil}^1 M$ is torsion free. The double root λ of $x^2 - a_p x + \varepsilon(p)p$ is in O_K, since it is integral and 2λ is in K. In the endomorphism ring of $A_{\mathbb{F}_p}$ we have the Eichler-Shimura congruence relation:

$$0 = (\mathrm{Frob}_p - \mathrm{Frob}_p)(\mathrm{Frob}_p - \mathrm{Frob}'_p) = \mathrm{Frob}_p^2 - a_p \mathrm{Frob}_p + \varepsilon(p)p = (\mathrm{Frob}_p - \lambda)^2, \tag{2.3}$$

where Frob_p is the Frobenius endomorphism and Frob'_p the Verschiebung, multiplied by $\varepsilon(p)$. The fact that every abelian variety over $\mathbb{F}_p$ is isogeneous to a product of simple ones implies that Frob_p is semi-simple in the sense that it satisfies an identity of the form $P(\mathrm{Frob}_p) = 0$, with P a polynomial with coefficients in $\mathbb{Q}$ that has simple roots. It follows that $\mathrm{Frob}_p = \lambda$ in $\mathrm{End}(A_{\mathbb{F}_p})$. Since $O_K \otimes \mathbb{Z}_p$ is a product of a finite number of discrete valuation rings, $\mathrm{Fil}^1 M$ is a locally free module over it; it is in fact free of rank one (consider $\mathbb{Q} \otimes \mathrm{Fil}^1 M$). It follows that λ does not annihilate $\mathbb{F}_p \otimes \mathrm{Fil}^1 M$, since we have $\lambda^2 = \varepsilon(p)p$. But $\mathbb{F}_p \otimes \mathrm{Fil}^1 M$ is the same as $\mathrm{H}^0(A_{\mathbb{F}_p}, \Omega^1)$, and on this module λ acts as Frob_p^*, hence it does annihilate. This contradiction finishes the proof. □

3 The general case.

In this section we try to generalize Theorem 2.1 as much as we can to higher weights. For doing that we replace the module M of Section 2 by the p-adic crystalline realization of the motive associated to f; this gives us a filtered ϕ-module M of rank two. The comparison theorem for crystalline and p-adic etale cohomology implies that this filtered ϕ-module is weakly admissible, from which it follows immediately that the crystalline Frobenius ϕ cannot be a scalar. Unfortunately, it is not known that ϕ is semi-simple, so all we show is that semi-simplicity of ϕ implies that the polynomial $x^2 - a_p x + \varepsilon(p)p^{k-1}$ has simple roots.

Let $f = \sum a_n q^n$ be a normalized cuspidal newform of some level N, weight $k \geq 2$ and character ε. Let K be the field generated by the a_n and the $\varepsilon(a)$. Let p be a prime number not dividing N. Our first objective is to construct the p-adic crystalline realization of the motive associated to f. In [14], Scholl constructs a Grothendieck motive $M(f)$ over $\mathbb{Q}$, with coefficients in K, such that for every prime number l the Galois representation $\rho_{f,l}: G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Q}_l \otimes K)$ is the dual of the l-adic realization $\mathrm{H}_l(M(f))$. Concretely, he constructs a projector in the group ring of a finite group of automorphisms of the smooth and projective model X of the $k-2$-fold fibered product of the universal elliptic curve over $Y(N')$ (with N' a suitable multiple of N) constructed by Deligne in [3, §5], such that the l-adic and Betti realizations of the corresponding Chow motive are, in a way that is compatible with Hecke operators, the parabolic cohomology groups used in [3]. The Grothendieck motive associated to this Chow motive (i.e., one replaces rational equivalence by homological equivalence) has an action by the Hecke algebra of the space of cuspidal modular forms of weight k on the modular curve $X(N')_{\mathbb{Q}}$; $M(f)$ is a suitable factor. The variety X over $\mathbb{Q}$ has a smooth projective model over $\mathbb{Z}[1/N']$ (see [14, 4.2.1] and [3]), hence $M(f)$ has a crystalline realization $M := \mathrm{H}_{\mathrm{crys}}(M(f))$ which is a free $\mathbb{Q}_p \otimes K$-module of rank two equipped with an endomorphism ϕ, the crystalline Frobenius, that is induced by the Frobenius endomorphism of the reduction mod p of X. The characteristic polynomial of ϕ is $x^2 - a_p x + \varepsilon(p)p^{k-1}$; this can be shown in the same way as one can show it for the l-adic realizations, or one invokes a result of Katz and Messing (see [14, 4.2.3]).

3.1 Theorem. *Let $f = \sum a_n q^n$ be a normalized cuspidal newform of some level N, weight $k \geq 2$ and character ε. Let p be a prime number not dividing N. Then the Frobenius ϕ of the crystalline realization M of the motive $M(f)$ is not scalar, i.e., it is not in $\mathbb{Q}_p \otimes K$.*

Proof. The proof is by contradiction. We suppose that ϕ is an element, λ say, of $\mathbb{Q}_p \otimes K$. The comparison theorem for crystalline and de Rham cohomology for smooth proper $\mathbb{Z}_p$-schemes (see [8, §1.3]) gives us an isomorphism between M and $\mathbb{Q}_p \otimes \mathrm{H}_{\mathrm{DR}}(M(f))$, and hence a decreasing filtration (the Hodge filtration) Fil on M. So (M, ϕ, Fil) is an object of the category of filtered ϕ-modules. (A filtered ϕ-module is a finite dimensional $\mathbb{Q}_p$-vector space M with a decreasing, exhaustive and separating filtration Fil^i, $i \in \mathbb{Z}$, and an endomorphism ϕ; morphisms are linear maps respecting Fil and ϕ; see [8, §2.3].) The Hodge filtration on $\mathrm{H}_{\mathrm{DR}}(M(f))$ induces the Hodge decomposition of $\mathbb{C}\otimes\mathrm{H}_{\mathrm{B}}(M(f))$, which is of type $(k-1,0),(0,k-1)$, hence $\mathrm{Fil}^0(M)=M$, $\mathrm{Fil}^1(M)=\mathrm{Fil}^{k-1}(M)$ is free of rank one, and $\mathrm{Fil}^k(M)=0$. Fontaine has constructed Grothendieck's "mysterious functor" $\mathrm{D}_{\mathrm{crys}}$ from the category of finite dimension $\mathbb{Q}_p$-vector spaces with continuous $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$-action to the category of filtered ϕ-modules (see the introduction of [8]). It is a theorem of Faltings (see [8, §3.2]), of which a special case was proved earlier by Fontaine and Messing, that there is an isomorphism of filtered ϕ-modules between M and $\mathrm{D}_{\mathrm{crys}}(\mathrm{H}_p(M(f)))$. In fact, the theorem is stated for varieties, but since the isomorphism is compatible with the multiplicative structure and with cycle classes, it also works for Grothendieck motives.

The most important consequence of this theorem for us is that the filtered ϕ-module M is admissible, hence weakly admissible, in the sense of Fontaine, see [7, 4.4.6]. Recall that to a filtered ϕ-module M one associates two polygons: the Hodge polygon, depending only on the filtration, and the Newton polygon, depending only on ϕ. Weakly admissible means that for every subobject M' of M the Newton polygon lies above or on the Hodge polygon, and that the two polygons for M itself have the same end-point. An equivalent formulation is the following. For M a filtered ϕ-module let $t_N(M)$ be the p-adic valuation of the determinant of ϕ, and let $t_H(M)$ be the maximal i such that $\mathrm{Fil}^i(\det M)\neq 0$. Then M is weakly admissible if and only if firstly: $t_N(M)=t_H(M)$, and secondly: for all subobjects M' of M one has $t_H(M')\leq t_N(M')$.

Consider now our weakly admissible filtered ϕ-module M. Since ϕ is the element λ of $\mathbb{Q}_p\otimes K$, we have the subobject $M':=\mathrm{Fil}^{k-1}(M)$ of M (we give it the induced filtration). Then $t_H(M')=[K:\mathbb{Q}](k-1)$, whereas $t_N(M')=[K:\mathbb{Q}](k-1)/2$ (recall that M' is free of rank one over $\mathbb{Q}_p\otimes K$ and that $\lambda^2=\varepsilon(p)p^{k-1}$). Since $k\geq 2$, this contradicts the weak admissibility of M. □

3.2 Corollary. *Let $f=\sum a_nq^n$ be a normalized cuspidal eigenform of some level N, weight $k\geq 2$ and character ε. Let p be a prime number not dividing N and suppose that the crystalline Frobenius ϕ of the $\mathbb{Q}_p$-vector space $M(f)$ is semi-simple. Then the polynomial $x^2-a_px+\varepsilon(p)p^{k-1}$ has simple roots.*

Proof. The proof is by contradiction: we suppose that $\lambda\in K$ is a double root. As we have already said above, the polynomial in question is the characteristic polynomial of the endomorphism ϕ of the free rank two $\mathbb{Q}_p\otimes K$-module M. Hence it satisfies the identity $(\phi-\lambda)^2=0$. Now $\mathbb{Q}_p\otimes K$ is a finite product of fields, hence the semi-simplicity of ϕ implies that ϕ is multiplication by λ. But this contradicts Theorem 3.1. □

3.3 Remark. The first three lines of [11, §2] show that Tate's conjecture mentioned in Section 1 implies the semi-simplicity of l-adic and crystalline Frobenius elements of smooth projective varieties over finite fields. □

3.4 Remark. Scholl remarks that his explicit construction of the crystalline realization M of $M(f)$ in [15] should show directly that M is weakly admissible. □

4 Applications.

4.1 Theorem. *Let $N \geq 1$ and $k \geq 2$ be integers. Let $f = \sum a_n q^n$ be a normalized cuspidal eigenform of level N and weight k. Let p be a prime number not dividing N. If $k > 2$ assume Tate's conjecture mentioned in Section 1. Then we have $|a_p| < 2p^{(k-1)/2}$.*

Proof. Let ε be the character of f. Theorems 2.1 and 3.1 show that $x^2 - a_p x + \varepsilon(p)p^{k-1}$ has no double root. Deligne has shown ([5, Thm. 6.1] and [4, Thm. 1.6]) that the roots have absolute value $p^{(k-1)/2}$. □

4.2 Theorem. *Let $N \geq 1$ and $k \geq 2$ be integers. Let p be a prime number such that p^3 does not divide N. Assume Tate's conjecture mentioned in Section 1 if $k \geq 3$ and $p|N$. Then the endomorphism T_p of $M^0(N,k)_{\mathbb{C}}$ is semi-simple.*

Proof. For the sake of notation, let $N \geq 1$ be an integer and p a prime number not dividing N. Let $k \geq 1$. In this case T_p is normal with respect to the Petersson scalar product on $M^0(N,k)_{\mathbb{C}}$. Hence T_p is diagonalizable.

Let us now consider $M^0(pN,k)_{\mathbb{C}}$, with $k \geq 2$. By the theory of newforms, $M^0(pN,k)_{\mathbb{C}}$ is the direct sum of its p-new part $M^0(pN,k)_{\mathbb{C}}^{\text{pnew}}$ and its p-old part $M^0(pN,k)_{\mathbb{C}}^{\text{pold}}$, this decomposition being respected by all Hecke operators. The restriction of $T_p^{pN,k}$ to $M^0(pN,k)_{\mathbb{C}}^{\text{pnew}}$ is normal, hence diagonalizable. The p-old part is isomorphic to the direct sum of two copies of $M^0(N,k)_{\mathbb{C}}$, via the map $(f(q), g(q)) \mapsto f(q) + g(q^p)$. The restriction of $T_p^{pN,k}$ to $M^0(pN,k)_{\mathbb{C}}^{\text{pold}}$ is then given by the following two by two matrix:

$$T_p^{pN,k}|_{M^0(pN,k)_{\mathbb{C}}^{\text{pold}}} = \begin{pmatrix} T_p^{N,k} & 1 \\ -p^{k-1}\langle p \rangle & 0 \end{pmatrix}. \tag{4.2.1}$$

We have already seen that $M^0(N,k)_{\mathbb{C}}$ is the direct sum of its common eigenspaces $V_{a_p,\varepsilon}$ for T_p and the diamond operators. It follows that $M^0(pN,k)_{\mathbb{C}}^{\text{pold}}$ decomposes as a direct sum of terms $V_{a_p,\varepsilon}^2$, and that the restriction of $T_p^{pN,k}$ to each of the $V_{a_p,\varepsilon}^2$ is annihilated by $x^2 - a_p x + \varepsilon(p)p^{k-1}$. Under the hypotheses of the theorem we are proving, these polynomials have simple roots by Theorems 2.1 and 3.1.

Let us now consider $M^0(p^2N,k)_{\mathbb{C}}$, with $k \geq 2$. Here too this space is the direct sum of its p-old and p-new parts. On the p-new part $T_p^{p^2N,k}$ is self-adjoint, hence diagonalizable. The p-old part is now isomorphic to the direct sum of three copies of $M^0(N,k)_{\mathbb{C}}$ and two copies of $M^0(pN,k)_{\mathbb{C}}^{\text{pnew}}$. The restrictions of $T_p^{p^2N,k}$ to $M^0(N,k)_{\mathbb{C}}^3$ and $(M^0(pN,k)_{\mathbb{C}}^{\text{pnew}})^2$ are given by the following matrices:

$$\begin{pmatrix} T_p^{N,k} & 1 & 0 \\ -p^{k-1}\langle p \rangle & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} T_p^{pN,k} & 1 \\ 0 & 0 \end{pmatrix} \tag{4.2.2}$$

One can now repeat the same type of argument as above, invoking Theorems 2.1 and 3.1 to see that, under the hypotheses of the theorem we are proving, $x(x^2 - a_p x + \varepsilon(p)p^{k-1})$, with $(a_p, \varepsilon(p))$ as before, has simple roots. The space $M^0(pN,k)_{\mathbb{C}}^{\text{pnew}}$ is a direct sum of eigenspaces for $T_p^{Np,k}$, and one knows that the eigenvalues are non-zero (see [5, §1.8]). It follows that also the restriction of $T_p^{p^2N,k}$ to $(M^0(pN,k)_{\mathbb{C}}^{\text{pnew}})^2$ is diagonalizable. □

4.3 Corollary. *Let $N \geq 1$ be cube free, and let $k \geq 2$. Let $\mathbb{T}$ be the $\mathbb{Z}$-algebra generated by the endomorphisms T_n, $n \geq 1$, and $\langle a \rangle$, $a \in (\mathbb{Z}/N\mathbb{Z})^*$, of $M^0(N,k)_\mathbb{C}$. Assume Tate's conjecture mentioned in Section 1 if $k > 2$. Then the ring $\mathbb{T}$ is reduced.*

Proof. This is so because $\mathbb{T}$ is a subring of the $\mathbb{C}$-algebra $\mathbb{T}_\mathbb{C} := \mathbb{C} \otimes \mathbb{T}$ generated by the T_n and $\langle a \rangle$. Theorem 4.2 tells us that the T_p and $\langle a \rangle$ can be simultaneously diagonalized. They generate $\mathbb{T}_\mathbb{C}$, hence $\mathbb{T}_\mathbb{C}$ is a product of copies of $\mathbb{C}$, hence reduced. □

4.4 Remark. For general N and k, the Hecke algebra $\mathbb{T}$ is well known to be a free $\mathbb{Z}$-module; $\mathbb{T}_\mathbb{Q} := \mathbb{Q} \otimes \mathbb{T}$ is well known to be Gorenstein, i.e., its $\mathbb{Q}$-linear dual $(\mathbb{Q} \otimes \mathbb{T})^\vee$ is free of rank one as $\mathbb{Q} \otimes \mathbb{T}$-module, see for example [18, p. 481]. One way to prove this is as follows. By the q-expansion principle, $M^0(N,k)_\mathbb{C}^\vee$ is free of rank one as $\mathbb{T}_\mathbb{C}$-module. Then one constructs a $\mathbb{T}_\mathbb{C}$-bilinear $\mathbb{C}$-valued pairing on $M^0(N,k)_\mathbb{C}$ to get an isomorphism of $\mathbb{T}_\mathbb{C}$-modules between $M^0(N,k)_\mathbb{C}$ and its dual. Another way to prove it is to use the theory of new forms. This last proof gives more information on how exactly $\mathbb{T}_\mathbb{C}$ decomposes as a product of $\mathbb{C}$-algebras. Parent needed such information and so he worked out the details in [13]. His work made see that Theorem 4.2 should be stated for N cube free instead of square free. Of course, statements that completions of $\mathbb{T}$ at certain of its maximal ideals are Gorenstein are much more subtle and harder to prove (see for example [18, §2.1]). □

5 Discriminants of Hecke algebras.

According to Corollary 4.3, certain Hecke algebras $\mathbb{T}$ are reduced. This means that the discriminants $\mathrm{discr}(\mathbb{T})$ of their trace forms $(x,y) \mapsto \mathrm{trace}(xy)$ are non-zero. These discriminants "count" all congruences between different eigenforms of fixed level and weight, hence are quite useless for dealing with congruences with a fixed form (in particular, nothing interesting can be said on the degrees of modular parametrizations of elliptic curves over $\mathbb{Q}$). The following result, relating such discriminants to heights of modular curves, is due to Abbes and Ullmo (unpublished).

5.1 Theorem. (Abbes, Ullmo) *Let p be a prime number, and let $\mathbb{T}$ be the Hecke algebra associated to $M^0(p,2)_\mathbb{C} = \mathrm{H}^0(X_0(p)_\mathbb{C}, \Omega)$. Then one has:*

$$h(X_0(p)_\mathbb{Q}) = \frac{1}{2}\log|\mathrm{discr}(\mathbb{T})| - \sum_{i=1}^{g} \log\|\omega_i\|,$$

where h is the modular height of curves over $\mathbb{Q}$ (see [16, §3.3]), where $\omega_1, \ldots, \omega_g$ are the normalized eigenforms and $\|\cdot\|$ the norm of the scalar product $\langle\omega|\eta\rangle = (i/2)\int_{X_0(p)(\mathbb{C})} \omega \wedge \overline{\eta}$ on $\mathrm{H}^0(X_0(p)_\mathbb{C}, \Omega)$.

Proof. We start by recalling the definition of h. So let $X_\mathbb{Q}$ be a smooth proper geometrically irreducible curve over $\mathbb{Q}$, of some genus g. Let $J_\mathbb{Q}$ be its jacobian and J the Néron model over $\mathbb{Z}$. Then we have the free $\mathbb{Z}$-module of rank one $\omega_J := \Lambda^g 0^* \Omega^1_{J/\mathbb{Z}}$, with the scalar product on $\mathbb{C} \otimes \omega_J$ given by $\langle\omega|\eta\rangle = (i/2)^g(-1)^{g(g-1)/2}\int_{J(\mathbb{C})} \omega \wedge \overline{\eta}$. The height $h(X_\mathbb{Q})$ is then defined to be the Arakelov degree of this metrized line bundle: $h(X_\mathbb{Q}) = \deg_{\mathrm{Ar}}(\omega_J) = -\log\|\omega\|$, with ω a generator of ω_J. The $\mathbb{Z}$-module ω_J is equipped with the scalar product on $\mathbb{C} \otimes \omega_J = \mathrm{H}^0(X_0(p)_\mathbb{C}, \Omega)$ already mentioned in the theorem above. This scalar product induces a real scalar product, and hence a volume form, on $\mathbb{R}\otimes\omega_J$ (the volume form being determined by the condition that a cube with edges of length one has volume one). A calculation (see [16, lemme 3.2.1]) shows that one has:

$$h(X_{\mathbb{Q}}) = -\log \mathrm{vol}(\mathbb{R}\otimes\omega_J/\omega_J). \tag{5.1.1}$$

Let now X be the curve $X_0(p)_{\mathbb{Q}}$. In that case, ω_J is the same as $\mathrm{H}^0(X_{\mathbb{Z}},\Omega)$, with $X_{\mathbb{Z}}$ the usual model over $\mathbb{Z}$ (semi-stable, $X_{\mathbb{F}_p}$ consisting of two irreducible components) and Ω its dualizing sheaf ([10, Ch. II, §3]). The fact that the two irreducible components of $X_{\mathbb{F}_p}$ are of genus zero implies that the pairing

$$\mathbb{T}\times\omega_J\to\mathbb{Z},\quad (t,\omega)\mapsto a_1(t\omega), \tag{5.1.2}$$

with a_1 denoting the linear form on ω_J that takes the coefficient of q in the q-expansion, is perfect, i.e., it induces an isomorphism between $\mathbb{T}^\vee$ of $\mathbb{T}$ and ω_J. Let $\omega_1,\ldots,\omega_g$ be as in the theorem. Sending an element of $\mathbb{T}_{\mathbb{R}}$ to the eigenvalues of the ω_i for it is an isomorphism of $\mathbb{R}$-algebras:

$$\mathbb{T}_{\mathbb{R}}\to\mathbb{R}^g,\quad t\mapsto (a_1(t\omega_1),\ldots,a_1(t\omega_g)). \tag{5.1.3}$$

The trace form on $\mathbb{T}$ corresponds to the standard scalar product on $\mathbb{R}^g$. Composing the dual of the isomorphism (5.1.3) with the isomorphism $\mathbb{T}_{\mathbb{R}}^\vee\to\mathbb{R}\otimes\omega_J$ from (5.1.2) gives an isomorphism from $\mathbb{R}^g$ to $\mathbb{R}\otimes\omega_J$ mapping the ith standard basis vector e_i to ω_i. It follows that the volume form on $\mathbb{R}\otimes\omega_J$ corresponds to $\prod_i\|\omega_i\|$ times the one on $\mathbb{T}_{\mathbb{R}}^\vee$ corresponding to the trace form. We find:

$$\mathrm{vol}(\mathbb{R}\otimes\omega_J/\omega_J) = \left(\prod_{i=1}^g\|\omega_i\|\right)\mathrm{vol}(\mathbb{T}_{\mathbb{R}}^\vee/\mathbb{T}) = \left(\prod_{i=1}^g\|\omega_i\|\right)\mathrm{vol}(\mathbb{T}_{\mathbb{R}}/\mathbb{T})^{-1}. \tag{5.1.4}$$

The proof is finished since $\mathrm{vol}(\mathbb{T}_{\mathbb{R}}/\mathbb{T}) = |\mathrm{discr}(\mathbb{T})|^{1/2}$. □

5.2 Theorem. (Abbes, Ullmo) *For every $\varepsilon>0$ there exists $c(\varepsilon)$ in $\mathbb{R}$ such that for all prime numbers p one has $h(X_0(p)_{\mathbb{Q}})\leq c(\varepsilon)p^{1+\varepsilon}$.*

Proof. Let $\mathbb{T}$ and g be as above, and let $\mathbb{T}' := \sum_{i=1}^g\mathbb{Z}T_i$. Then $\mathbb{T}'$ is of finite index in $\mathbb{T}$ because ∞ is not a Weierstrass point of $X_0(p)$ (see [9, §3] or [6, §4]). The image of T_i under the isomorphism (5.1.3) is $(a_i(\omega_1),\ldots,a_i(\omega_g))$. It follows that we have the equalities:

$$\mathrm{discr}(\mathbb{T}) = \mathrm{discr}(\mathbb{T}')|\mathbb{T}/\mathbb{T}'|^{-2},\quad |\mathrm{discr}(\mathbb{T}')| = |\det_{1\leq i,j\leq g} a_i(\omega_j)|^2. \tag{5.2.1}$$

Weil's theorem on absolute values of eigenvalues of Frobenius endomorphisms of abelian varieties over finite fields implies that $|a_i(\omega_j)|\leq\sigma(i)i^{1/2}$, where $\sigma(i)$ is the number of positive integers dividing i. It follows that $|\mathrm{discr}(\mathbb{T}')|\leq g!\prod_{i=1}^g\sigma(i)i^{1/2}$. The rest of the proof consists of applying Theorem 5.1 and standard estimates (including an absolute lower bound for the $\|\omega_i\|$). □

5.3 Remark. One knows (see [10, Ch. II, Prop. 10.6]) that $\mathrm{Spec}(\mathbb{T})$ (for $\mathbb{T}$ as above) is connected. This implies a lower bound for $\mathrm{discr}(\mathbb{T})$ (use [12]). On the other hand, the $\|\omega_i\|$ are bounded above by a constant times p. Unfortunately, the lower bound for $h(X_0(p)_{\mathbb{Q}})$ obtained like this seems too weak to be useful. Assuming $\mathbb{T}$ to be Gorenstein does not significantly improve this lower bound. □

5.4 Remark. Several problems arise when one wants to generalize the above results for $X_0(p)_{\mathbb{Q}}$ to more general $X_0(N)_{\mathbb{Q}}$. First of all, $\mathbb{T}^\vee$ will not be the same as ω_J, but it should be possible to estimate $|\mathbb{T}^\vee/\omega_J|$. Secondly, the comparison of the trace form on $\mathbb{T}$ and the scalar product on $\mathbb{R}\otimes\omega_J$ is more complicated. Note that the trace form can even be degenerate, if N is not cube free. Thirdly, it seems to be unknown if ∞ can be a Weierstrass point on $X_0(N)_{\mathbb{Q}}$ when N is square free (see [9, §3]; it is known that ∞ is a Weierstrass point when N is divisible by 4 or 9, for example).

Suppose now that N is square free. Before knowing the result that $\mathbb{T}$ is reduced, Abbes and Ullmo have related $h(X_0(N))$ to the discriminants of the "new parts" of the Hecke algebras of level dividing N (unpublished). The techniques they use come from [1] and [2]. They show that Theorem 5.2 holds for square free N such that ∞ is not a Weierstrass point. With the same techniques it is certainly possible to solve the first two of the three problems mentioned above. □

Acknowledgements. We thank Rutger Noot for a discussion during which we were led to a proof of Theorem 2.1. Bas Edixhoven is grateful to Richard Taylor for a discussion concerning the case of general weights. He also thanks Pierre Berthelot for explaining to him certain results on Newton and Hodge polygons. Thanks also go to Ahmed Abbes and Emmanuel Ullmo for communicating to us their results mentioned in Section 5, and permitting us to publish them. Bas Edixhoven wants to thank the Centre for Research in Mathematics at the Institut d'Estudis Catalans in Barcelona where he finally found time to write this text. Robert Coleman thanks the University of Rennes for inviting him for a month in 1995.

References

[1] A. Abbes and E. Ullmo. *Comparaison des métriques d'Arakelov et de poincaré sur* $X_0(N)$. To appear in Duke Math. Journal.

[2] A. Abbes and E. Ullmo. *A propos de la conjecture de Manin pour les courbes elliptiques modulaires.* To appear in Compositio Mathematica.

[3] P. Deligne. *Formes modulaires et representations* l*-adiques.* Séminaire Bourbaki exp. 355. Lecture Notes in Mathematics **179**, Springer, Heidelberg, 1969.

[4] P. Deligne. *La conjecture de Weil, I.* Publ. Math. I.H.E.S. **43**, 1974, 273–307.

[5] P. Deligne and J-P. Serre. *Formes modulaires de poids 1.* Ann. Sci. Ec. Norm. Sup. **7** (1974), 507–530. Also: J-P. Serre, Collected papers, 101.

[6] S.J. Edixhoven. *Rational torsion points on elliptic curves over number fields, after Kamienny and Mazur.* Astérisque **227** (1995), 209–227.

[7] J-M. Fontaine. *Représentations* p*-adiques semi-stables.* Exposé III in Astérisque **223**, Périodes p-adiques, 1994.

[8] L. Illusie. *Crystalline cohomology.* Proceedings of symposia in pure mathematics, Volume **55**, Part 1.

[9] J. Lehner and M. Newman. *Weierstrass points of* $\Gamma_0(N)$. Annals of Math. **79** (1963), no. 2, 360–368.

[10] B. Mazur. *Modular curves and the Eisenstein ideal.* Publ. Math. I.H.E.S. **47** (1977), 33–186.

[11] J.H. Milne. *Motives over finite fields.* Proceedings of symposia in pure mathematics, Volume **55**, Part 1, 401–459.

[12] A. Odlyzko. *Lower bounds for discriminants of number fields.* Acta Arithmetica **29** (1976), 275–296.

[13] P. Parent. Article in preparation.

[14] A.J. Scholl. *Motives for modular forms.* Invent. math. **100** (1990), 419–430.

[15] A.J. Scholl. *Modular forms and de Rham cohomology; Atkin-Swinnerton-Dyer congruences.* Invent. math. **79** (1985), 49–77.

[16] L. Szpiro. *Degrés, intersections, hauteurs.* Exposé I in Astérisque **127** (1985).

[17] D. Ulmer. *A construction on local points on elliptic curves over local fields.* International Mathematics Research Notices No. **7** (1995).

[18] A. Wiles. *Modular elliptic curves and Fermat's Last Theorem..* Annals of Mathematics **142** (1995), 443–551.

Robert Coleman
Dept. of Mathematics
University of California
Berkeley, CA 94720
U.S.A.

Bas Edixhoven
IRMAR
Campus de Beaulieu
35042 Rennes cedex
France