# A RESULT ON MODULAR FORMS IN CHARACTERISTIC p

Nicholas M. Katz

ABSTRACT

The action of the derivation $\theta = q\frac{d}{dq}$ on the q-expansions of modular forms in characteristic p is one of the fundamental tools in the Serre/Swinnerton-Dyer theory of mod p modular forms. In this note, we extend the basic results about this action, already known for $p \geq 5$ and level one, to arbitrary p and arbitrary prime-to-p level.

## I. Review of modular forms in characteristic p

We fix an algebraically closed field K of characteristic $p > 0$, an integer $N \geq 3$ prime to p, and a primitive N'th root of unity $\zeta \in K$. The moduli problem "elliptic curves E over K-algebras with level N structure $\alpha$ of determinant $\zeta$" is represented by

$$\begin{array}{c}(E_{univ}, \alpha_{univ}) \\ \big\downarrow{\scriptstyle \pi} \\ M_N\end{array}$$

with $M_N$ a smooth affine irreducible curve over K. In terms of the invertible sheaf on $M_N$

$$\underline{\omega} = \pi_* \Omega^1_{E_{univ}/M_N} \quad ,$$

the graded ring $R_N^{\cdot}$ of (not necessarily holomorphic at the cusps) level N modular forms over K is

$$\bigoplus_{k\in\mathbf{Z}} H^0(M_N, \underline{\omega}^{\otimes k})$$

Given a K-algebra B, a test object $(E,\alpha)$ over B, and a nowhere-vanishing invariant differential $\omega$ on E, any element

$f \in R_N^{\cdot}$ (not necessarily homogeneous) has a value $f(E,\omega,\alpha) \in B$, and $f$ is determined by all of its values (cf. [2]).

Over $B = K((q^{1/N}))$, we have the Tate curve Tate(q) with its "canonical" differential $\omega_{can}$ (viewing Tate(q) as $\mathbb{G}_m/q^{\mathbb{Z}}$, $\omega_{can}$ "is" $dt/t$ from $\mathbb{G}_m$). By evaluating at the level $N$ structures $\alpha_0$ of determinant $\zeta$ on Tate(q), all of which are defined over $K((q^{1/n}))$, we obtain the q-expansions of elements $f \in R_N^{\cdot}$ at the corresponding cusps:

$$f_{\alpha_0}(q) \overset{dfn}{=\!=} f(\text{Tate}(q),\omega_{can},\alpha_0) \in K((q^{1/N})) \quad .$$

A homogeneous element $f \in R_N^k$ is uniquely determined by its weight $k$ and by any one of its q-expansions. A form $f \in R_N^k$ is said to be holomorphic if all of its q-expansions lie in $K[[q^{1/N}]]$, and to be a cusp form if all of its q-expansions lie in $q^{1/N}K[[q^{1/N}]]$. The holomorphic forms constitute a subring $R_{N,holo}^{\cdot}$ of $R_N^{\cdot}$, and the cusp forms are a graded ideal in $R_{N,holo}^{\cdot}$.

The Hasse invariant $A \in R_{N,holo}^{p-1}$ is defined modularly as follows. Given $(E,\omega,\alpha)$ over $B$, let $\eta \in H^1(E,\mathcal{O}_E)$ be the basis dual to $\omega \in H^0(E,\Omega^1_{E/B})$. The p'th power endomorphism $x \to x^p$ of $\mathcal{O}_E$ induces an endormorphism of $H^1(E,\mathcal{O}_E)$, which must carry $\eta$ to a multiple of itself. So we can write

$$\eta^p = A(E,\omega,\alpha)\cdot\eta \quad \text{in} \quad H^1(E,\mathcal{O}_E),$$

for some $A(E,\omega,\alpha) \in B$, which is the value of $A$ on $(E,\omega,\alpha)$. <u>All</u> the q-expansions of the Hasse invariant are identically 1:

$$A_{\alpha_0}(q) = 1 \quad \text{in} \quad K((q^{1/N})) \text{ for each } \alpha_0.$$

For each level $N$ structure $\alpha_0$ on Tate(q), the corresponding q-expansion defines ring homomorphisms

$$R_N^{\cdot} \longrightarrow K((q^{1/N})), \qquad R_{N,holo}^{\cdot} \longrightarrow K[[q^{1/N}]]$$

whose kernels are precisely the principal ideals $(A-1)R_N^{\cdot}$ and

$(A-1)R^{\cdot}_{N,\mathrm{holo}}$ respectively ([4], [5]).

A form $f \in R^k_N$ is said to be of <u>exact filtration</u> $k$ if it is not divisible by $A$ in $R^{\cdot}_N$, or equivalently, if there is no form $f' \in R^{k'}_N$ with $k' < k$ which, at some cusp, has the same q-expansion that $f$ does.

## II. Statment of the theorem, and its corollaries

The following theorem is due to Serre and Swinnerton-Dyer ([4], [5]) in characteristic $p \geq 5$, and level $N = 1$.

<u>Theorem.</u>

(1) There exists a derivation $A\theta : R^{\cdot}_N \to R^{\cdot+p+1}_N$ which increases degrees by $p+1$, and whose effect upon each q-expansion is $q\frac{d}{dq}$:

$$(A\theta f)_{\alpha_0}(q) = q\frac{d}{dq}(f_{\alpha_0}(q)) \text{ for each } \alpha_0 .$$

(2) If $f \in R^k_N$ has exact filtration $k$, and $p$ does not divide $k$, then $A\theta f$ has exact filtration $k+p+1$, and in particular $A\theta f \neq 0$.

(3) If $f \in R^{pk}_N$ and $A\theta f = 0$, then $f = g^p$ for a unique $g \in R^k_N$.

<u>Some Corollaries</u>

(1) The operator $A\theta$ maps the subring of holomorphic forms to the ideal of cusp forms. (Look at q-expansions.)

(2) If $f$ is non-zero and holomorphic, of weight $1 \leq k \leq p-2$, then $f$ has exact filtration $k$. (For if $f = Ag$, then $g$ is holomorphic of weight $k-(p-1) < 0$, hence $g = 0$.)

(3) If $1 \leq k \leq p-2$, the map $A\theta : R^k_{N,\mathrm{holo}} \to R^{k+p+1}_{N,\mathrm{holo}}$ is injective. (This follows from (2) above and the theorem.)

(4) If $f$ is non-zero and holomorphic of weight $p-1$, and vanishes at some cusp, then $f$ has exact filtration $p-1$. (For if $f = Ag$, then $g$ is holomorphic of weight 0, hence constant; as $g$ vanishes at one cusp, it must be zero.)

(5) (determination of $\mathrm{Ker}(A\theta)$). If $f \in R_N^k$ has $A\theta f = 0$, then we can uniquely write $f = A^r \cdot g^p$ with $0 \leq r \leq p-1$, $r + k \equiv 0 (\mathrm{mod}\ p)$, and $g \in R_N^\ell$ with $p\ell + r(p-1) = k$. (This is proven by induction on $r$, the case $r = 0$ being part (3) of the theorem. If $r \neq 0$, then $k \not\equiv 0(p)$, but $A\theta f = 0$. Hence by part (2) of the theorem $f = Ah$ for some $h \in R_N^{k+1-p}$. Because $f$ and $h$ have the same q-expansions, we have $A\theta h = 0$, and $h$ has lower $r$.)

(6) In (5) above, if $f$ is holomorphic (resp. a cusp form, resp. invariant by a subgroup of $SL_2(Z/NZ)$), so is $g$ (by unicity of $g$).

III. Beginning of the proof: defining $\theta$ and $A\theta$, and proving part (1)

The absolute Frobenius endomorphism $F$ of $M_N$ induces an F-linear endomorphism of $H^1_{DR}(E_{univ}/M_N)$, as follows. The pull-back $E_{univ}^{(F)}$ of $E_{univ}$ is obtained by dividing $E_{univ}$ by its finite flat rank $p$ subgroup scheme $\mathrm{Ker}\ Fr$ where

$$Fr: E_{univ} \longrightarrow E_{univ}^{(F)}$$

is the relative Frobenius morphism. The desired map is $Fr^*$

$$\begin{array}{ccc} Fr^*: H^1_{DR}(E_{univ}^{(F)}/M_N) & \longrightarrow & H^1_{DR}(E_{univ}/M_N) \\ \uparrow \wr & & \\ H^1_{DR}(E_{univ}/M_N)^{(F)} & & \end{array}$$

Lemma 1. The image $U$ of $Fr^*$ is a locally free submodule of $H^1_{DR}(E_{univ}/M_N)$ of rank one, with the quotient $H^1_{DR}(E_{univ}/M_N)/U$ locally free of rank one. The open set $M_N^{Hasse} \subset M_N$ where $A$ is invertible is the largest open set over which $U$ splits the Hodge filtration, i.e., where $\underline{\omega} \oplus U \xrightarrow{\sim} H^1_{DR}(E_{univ}/M_N)$.

Proof. Because $Fr^*$ kills $H^0(E_{univ}, \Omega^1 E_{univ/M_N})^{(F)}$ it factors through the quotient $H^1(E_{univ}, \mathcal{O})^{(F)}$, where it induces the inclusion map in the "conjugate filtration" short exact sequence (cf [1], 2.3)

$$0 \longrightarrow H^1(E_{univ}, \mathcal{O})^{(F)} \xrightarrow{Fr^*} H^1_{DR}(E_{univ}/M_N) \longrightarrow H^0(E_{univ}, \Omega^1 E_{univ/M_N}) \longrightarrow 0.$$

This proves the first part of the lemma. To see where $U$ splits the Hodge filtration, we can work locally on $M_N$. Choose a basis $\omega, \eta$ of $H^1_{DR}$ adapted to the Hodge filtration, and satisfying $\langle\omega,\eta\rangle_{DR} = 1$. Then $\eta$ projects to a basis of $H^1(E,\mathcal{O}_E)$ dual to $\omega$, and so the matrix of $Fr^*$ on $H^1_{DR}$ is (remembering $Fr^*(\omega^{(F)}) = 0$)

$$\begin{pmatrix} 0 & B \\ 0 & A \end{pmatrix}$$

where $A$ is the value of the Hasse invariant. Thus $U$ is spanned by $B\omega + A\eta$, and the condition that $\omega$ and $B\omega + A\eta$ together span $H^1_{DR}$ is precisely that $A$ be invertible. Q.E.D.

Remark. According to the first part of the lemma, the functions $A$ and $B$ which occur in the above matrix have no common zero. This will be crucial later.

We can now define a derivation $\theta$ of $R_N[1/A]$ as follows. (Compare [2], A1.4.) Over $M_N^{Hasse}$, we have the decomposition

$$H^1_{DR} \simeq \underline{\omega} \oplus U \quad ,$$

which for each integer $k \geq 1$ induces a decomposition

$$\mathrm{Symm}^k H^1_{DR} \simeq \underline{\omega}^{\otimes k} \oplus (\underline{\omega}^{\otimes k-1} \otimes U) \oplus \ldots \oplus U^{\otimes k}$$

The Gauss-Manin connection

$$\nabla : H^1_{DR} \longrightarrow H^1_{DR} \otimes \Omega^1_{M_N/K}$$

induces, for each $k \geq 1$ a connection

$$\nabla : \mathrm{Symm}^k H^1_{DR} \longrightarrow (\mathrm{Symm}^k H^1_{DR}) \otimes \Omega^1_{M_N/K} \quad .$$

Using the Kodaira-Spencer isomorphism ([2], A.1.3.17)

$$\underline{\omega}^{\otimes 2} \xrightarrow{\sim} \Omega^1_{M_N/K}$$

we can define a mapping of sheaves

$$\theta : \underline{\omega}^{\otimes k} \longrightarrow \underline{\omega}^{\otimes k+2}$$

as the composite

$$\begin{array}{ccl}
\underline{\omega}^{\otimes k} & \hookrightarrow & \mathrm{Symm}^k H^1_{DR} = \underline{\omega}^{\otimes k} \oplus \ldots \\
 & & \downarrow \nabla \\
\theta & & (\mathrm{Symm}^k H^1_{DR}) \otimes \Omega^1_{M_N} \\
 & & \downarrow \int \mathrm{KS} \\
 & & (\mathrm{Symm}^k H^1_{DR}) \otimes \underline{\omega}^{\otimes 2} \simeq \underline{\omega}^{\otimes k+2} \oplus \ldots \\
 & & \qquad\qquad \downarrow \mathrm{pr}_1 \\
 & \searrow & \qquad\qquad \underline{\omega}^{\otimes k+2}
\end{array}$$

Passing to global sections over $M_N^{\mathrm{Hasse}}$, we obtain, for $k \geq 1$, a map

$$\theta : H^0(M_N^{\mathrm{Hasse}}, \underline{\omega}^{\otimes k}) \longrightarrow H^0(M_N^{\mathrm{Hasse}}, \underline{\omega}^{\otimes k+2}) \quad .$$

<u>Lemma 2.</u> The effect of $\theta$ upon q-expansions is $q \frac{d}{dq}$ .

<u>Proof</u>. Consider Tate(q) with its canonical differential $\omega_{\mathrm{can}}^{\otimes 2}$ over $k((q^{1/N}))$ Under the Kodaira-Spencer isomorphism, $\omega_{\mathrm{can}}$ corresponds to $dq/q$, the dual derivation to which is $q \frac{d}{dq}$ . By the explicit calculations of ([2], A.2.2.7), U is spanned by $\nabla(q \frac{d}{dq})(\omega_{\mathrm{can}})$. Thus given an element $f \in H^0(M_N^{\mathrm{Hasse}}, \underline{\omega}^{\otimes k})$, its local expression as a section of $\underline{\omega}^{\otimes k}$ on (Tate(q), some $\alpha_0$) is $f_{\alpha_0}(q) \cdot \omega_{\mathrm{can}}^{\otimes k}$. Thus

$$\begin{aligned}
\nabla(f_{\alpha_0}(q) \cdot \omega_{\mathrm{can}}^{\otimes k}) &= \nabla(q \tfrac{d}{dq})(f_{\alpha_0}(q) \cdot \omega_{\mathrm{can}}^{\otimes k}) \cdot \frac{dq}{q} \\
&= \nabla(q \tfrac{d}{dq})(f_{\alpha_0}(q) \cdot \omega_{\mathrm{can}}^{\otimes k}) \cdot \omega_{\mathrm{can}}^{\otimes 2} \\
&= q \tfrac{d}{dq} (f_{\alpha_0}(q)) \cdot \omega_{\mathrm{can}}^{\otimes k+2} + k \cdot f_{\alpha_0}(q) \cdot \omega_{\mathrm{can}}^{\otimes k+1} \cdot \nabla(q \tfrac{d}{dq})(\omega_{\mathrm{can}}).
\end{aligned}$$

Because $\nabla (q \frac{d}{dq}) (\omega_{\mathrm{can}})$ lies in U, it follows from the definition of $\theta$ that we have $(\theta f)_{\alpha_0}(q) = q\,\frac{d}{dq}\,(f_{\alpha_0}(q))$.

Q.E.D.

Lemma 3. For $k \geq 1$, there is a unique map $A\theta : R_N^k \longrightarrow R_N^{k+p+1}$ such that the diagram below commutes

$$\begin{array}{ccccc} H^0(M_N^{\text{Hasse}}, \underline{\omega}^{\otimes k}) & \xrightarrow{\theta} & H^0(M_N^{\text{Hasse}}, \underline{\omega}^{\otimes k+2}) & \xrightarrow{\times A} & H^0(M_N^{\text{Hasse}}, \underline{\omega}^{\otimes k+p+1}) \\ \cup & & & & \cup \\ R_N^k = H^0(M_N, \underline{\omega}^{\otimes k}) & & \xrightarrow{\quad A\theta \quad} & & R_N^{k+p+1} = H^0(M_N, \underline{\omega}^{\otimes k+p+1}) \end{array}$$

Proof. Again we work locally on $M_N$. Let $\omega$ be a local basis of $\underline{\omega}$, $\xi$ the local basis of $\Omega^1_{M_N/K}$ corresponding to by the Kodaira-Spencer isomorphism, $D$ the local basis of $\underline{\mathrm{Der}}_{M_N/K}$ dual to $\xi$, and $\omega' = \nabla(D)\omega \in H^1_{DR}$. Then $\langle\omega,\omega'\rangle_{DR} = 1$, (this characterizes $D$), so that $\omega$ and $\omega'$ form a basis of $H^1_{DR}$, adapted to the Hodge filtration, in terms of which the matrix of $\mathrm{Fr}^*$ is

$$\begin{pmatrix} 0 & B \\ 0 & A \end{pmatrix}$$

with $A = A(E,\omega)$. Let $u \in U$ be the basis of $U$ over $M_N^{\text{Hasse}}$ which is dual to $\omega$. Then $u$ is proportional to $B\omega + A\omega'$, and satisfies $\langle\omega,\omega'\rangle_{DR} = 1$, so that

$$u = \frac{B}{A}\,\omega + \omega' \ .$$

In terms of all this, we will compute $\theta f$ for $f \in R_N^k$, and show that it has at worst a single power of $A$ in its denominator. Locally, $f$ is the section $f_1 \cdot \omega^{\otimes k}$ of $\underline{\omega}^{\otimes k}$, with $f_1$ holomorphic.

$$\begin{aligned} \nabla(f_1\omega^{\otimes k}) &= \nabla(D)(f_1\omega^{\otimes k})\cdot\xi \\ &= \nabla(D)(f_1\omega^{\otimes k})\cdot\omega^{\otimes 2} \\ &= D(f_1)\cdot\omega^{\otimes k+2} + kf_1\omega^{\otimes k+1}\cdot\omega' \\ &= D(f_1)\omega^{\otimes k+2} + kf_1\omega^{\otimes k+1}(u - \tfrac{B}{A}\,\omega) \end{aligned}$$

$$= (D(f_1) - kf_1 \cdot \frac{B}{A})\, \omega^{\otimes k+2} + kf_1 \omega^{\otimes k+1} \cdot u \quad .$$

Thus from the definition of $\theta$ it follows that the local expression of $\theta(f)$ is

$$\theta(f) = (D(f_1) - kf_1 \frac{B}{A})\, \omega^{\otimes k+2} \quad . \qquad \text{Q.E.D.}$$

We can now conclude the proof of Part (1) of the theorem. Up to now, we have only defined A on elements of $R_N^{\cdot}$ of positive degree. But as $R_N^{\cdot}$ has units which are homogeneous of positive degree (e.g., $\Delta$), the derivation $A\theta$ extends uniquely to all of $R_N^{\cdot}$ by the explicit formula

$$A\theta f = \frac{A\theta(f\cdot\Delta^{pr})}{\Delta^{pr}} \quad \text{for} \quad r \gg 0 \quad .$$

The local expression for $A\theta(f)$

$$A\theta(f) = (AD(f_1) - kf_1 B)\omega^{\otimes k+p+1} \quad \text{for} \quad f \in R_N^k$$

remains valid.

IV. <u>Conclusion of the proof: Parts (2) and (3)</u>

Suppose $f \in R_N^k$ has exact filtration k. This means that f is not divisible by A in $R_N^{\cdot}$, i.e., that at some zero of A, f has a <u>lower order</u> zero (as section of $\underline{\omega}^{\otimes k}$) than A does (as section of $\underline{\omega}^{\otimes p-1}$). (In fact, we know by Igusa [3] that A has <u>simple</u> zeros, so in fact f must be <u>invertible</u> at some zero of A. Rather surprisingly, we will <u>not</u> make use of this fact.)

Locally on $M_N$, we pick a basis $\omega$ of $\underline{\omega}$. Then f becomes $f_1 \cdot \omega^{\otimes k}$, and $A\theta(f)$ is given by

$$A\theta(f_1 \cdot \omega^{\otimes k}) = (AD(f_1) - kBf_1)$$

Suppose now that k is not divisible by p. Recall that B is <u>invertible</u> at all zeros of A (cf the remark following Lemma 1). Thus if $x \in M_N$ is a zero of A where $\text{ord}_x(f_1) < \text{ord}_x(A)$, we easily compute

$$\mathrm{ord}_x(AD(f_1)-kBf_1) = \mathrm{ord}_x(f_1) < \mathrm{ord}_x(A).$$

This proves Part (2) of the theorem.

To prove Part (3), let $f \in R_N^{pk}$ have $A\theta(f) = 0$. The local expression for $A\theta(f)$ gives

$$AD(f_1)\omega^{\otimes k+p+1} = 0$$

and hence $D(f_1) = 0$. Because $M_N$ is a smooth curve over a perfect field of characteristic $p$, this implies that $f_1$ is a $p^{th}$ power, say $f_1 = (g_1)^p$. Thus $f_1\omega^{\otimes kp} = (g_1\omega^{\otimes k})^p$, so that $f$, as section of $\underline{\omega}^{\otimes kp}$, is, locally on $M_N$, the $p^{th}$ power of a (necessarily unique) section $g$ of $\underline{\omega}^{\otimes k}$. By unicity, these local $g$'s patch together. Q.E.D.

REFERENCES

[1] N. Katz, Algebraic solutions of differential equations, Inventiones Math. 18 (1972), 1-118.

[2] ———, P-adic properties of modular schemes and modular forms, Proceedings of the 1972 Antwerp International Summer School on Modular Functions, Springer Lecture Notes in Mathematics 350 (1973), 70-189.

[3] J. Igusa, Class number of a definite quaternion with prime discriminant, Proc. Nat'l. Acad. Sci. 44 (1958), 312-314.

[4] J. -P. Serre, Congruences et formes modulaires (d'apres H. P. F. Swinnerton-Dyer), Exposé 416, Séminaire N. Bourbaki, 1971/72, Springer Lecture Notes in Mathematics 317 (1973), 319-338.

[5] H. P. F. Swinnerton-Dyer, On $\ell$-adic representations and congruences for coefficients of modular forms, Proceedings of the 1972 Antwerp International Summer School on Modular Forms, Springer Lecture Notes in Mathematics 350 (1973), 1-55.

N. Katz
Departement of Mathematics
Fine Hall
Princeton University
Princeton, N.J. 08540