

On theta series mod p

By Masami OHTA

To the memory of Takuro Shintani

Introduction.

Let p be a fixed prime number, and let B denote the definite quaternion algebra over the rational number field \mathbf{Q} whose discriminant is p . If we denote by h the class number of B , we obtain h^2 theta series $\mathcal{G}_{i,j}$ ($1 \leq i, j \leq h$) from B , which are modular forms of weight 2 with respect to the group $\Gamma_0(p) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbf{Z}) \mid c \equiv 0 \pmod{p} \right\}$. Motivated by a conjecture of Hecke, Eichler [5] proved that $\mathcal{G}_{i,j}$ span the space $M_2(\Gamma_0(p))$ of modular forms of weight 2 with respect to $\Gamma_0(p)$. This "basis problem" has been generalized by Eichler himself, Hijikata-Saito and Pizer by generalizing the method of [5]. The purpose of this paper is *not* to go further in this direction, but just to give a new proof for the original result of Eichler [5] (which turns out to be simpler). We can in fact prove more:

THEOREM. *Suppose that $p \geq 5$. Then the coefficients of the q -expansions of $\mathcal{G}_{i,j}$ are p -integral, and their reductions mod p span the space of modular forms mod p of weight 2 with respect to $\Gamma_0(p)$, in the sense of Serre [11] and [12] (see the text for details).*

Note that, when $p=2$ or 3 , $M_2(\Gamma_0(p))$ is one dimensional, and hence the "basis problem" is trivial.

The content of this paper is as follows. In §1, after recalling the definitions of the Brandt matrices and the theta series, we give an interpretation of them in terms of supersingular elliptic curves in characteristic p . §2 is a preliminary section in which we recall known facts about modular forms mod p . Using a result of Atkin and Serre [1] (cf. Prop. 2), we prove, in §3, that the representation matrices of the Hecke operators acting on the space of modular forms mod p coincide with the Brandt matrices mod p (Prop. 3). Our main result will then follow easily from this.

The notation introduced here will be used throughout this paper.

§ 1. The Brandt matrices and the theta series.

1.1. The notation being as in the introduction, let us take a maximal order \mathfrak{o} of B , and a set of representatives $\alpha_1, \dots, \alpha_h$ of the left \mathfrak{o} -ideal classes. Let \mathfrak{o}_i be the right order of α_i , and denote by e_i the order of the unit group \mathfrak{o}_i^\times of \mathfrak{o}_i ($1 \leq i \leq h$). For a positive integer n , let $c_{ij}(n)$ ($1 \leq i, j \leq h$) denote the number of elements A of B such that $\alpha_i^{-1} \alpha_j A$ is integral and $N(\alpha_i^{-1} \alpha_j A) = n$, where N is the reduced norm of B over \mathbf{Q} . We put $b_{ij}(n) = c_{ij}(n) e_j^{-1}$ for i, j and n as above; $b_{ij}(n)$ is the number of integral left \mathfrak{o}_i -ideals which are left equivalent to $\alpha_i^{-1} \alpha_j$ and whose reduced norms are equal to n . We also put $c_{ij}(0) = 1$ and $b_{ij}(0) = e_j^{-1}$ ($1 \leq i, j \leq h$). Note that prime factors of e_i are at most 2 and 3, and hence $b_{ij}(n)$ are p -integral for any prime $p \geq 5$.

For a non-negative integer n , we denote by $B(n)$ (the Brandt matrix) the $h \times h$ matrix whose (i, j) -component is $b_{ij}(n)$; $B(n) = (b_{ij}(n))$. Let H be the complex upper half plane. For a variable z on H , we put $q = e^{2\pi iz}$.

DEFINITION. The notation being as above, we define the $h \times h$ matrix valued function $\Theta(z)$ on H by

$$\Theta(z) = \sum_{n=0}^{\infty} B(n) q^n.$$

Its (i, j) -component is denoted by $\mathcal{D}_{ij}(z)$; $\mathcal{D}_{ij}(z) = \sum_{n=0}^{\infty} b_{ij}(n) q^n$.

It is known that $\mathcal{D}_{ij}(z)$ are modular forms of weight 2 with respect to the group $\Gamma_0(p)$. But in the following, we forget about the analyticity and consider $\mathcal{D}_{ij}(z)$ and $\Theta(z)$ as formal power series in q with coefficients in $\mathbf{Z}[e_1^{-1}, \dots, e_h^{-1}]$, and write them \mathcal{D}_{ij} and Θ , respectively.

1.2. We next recall some facts about supersingular elliptic curves in characteristic p (cf. Deuring [3], Shimura, Taniyama [14]). Let \mathbf{F}_{p^n} be the finite field with p^n elements, and $\bar{\mathbf{F}}_p$ the algebraic closure of \mathbf{F}_p . It is known that there are exactly h non-isomorphic supersingular elliptic curves over $\bar{\mathbf{F}}_p$. Let j_1, \dots, j_h be their modular j -invariants. Then all the j_i are contained in \mathbf{F}_{p^2} . We take and fix an elliptic curve E_i defined over $\mathbf{F}_p(j_i)$ whose j -invariant is j_i ($1 \leq i \leq h$). Let $\text{End}(E_i)$ denote the ring of $\bar{\mathbf{F}}_p$ -endomorphisms of E_i , and put $\text{End}^0(E_i) = \text{End}(E_i) \otimes_{\mathbf{Z}} \mathbf{Q}$. For each i ($1 \leq i \leq h$), there is a ring isomorphism θ_i of B onto $\text{End}^0(E_i)$, and $\mathfrak{o}_i = \theta_i^{-1}(\text{End}(E_i))$ is a maximal order of B .

It is also known that, for each i , there is an integral left \mathfrak{o}_i -ideal α_i such that (E_i, θ_i) is an α_i -transform of (E_1, θ_1) in the sense of [14] 7.1. The left \mathfrak{o}_i -ideals $\alpha_1, \dots, \alpha_h$ constitute a set of representatives of the left \mathfrak{o} -ideal classes, and hence we may use these $\alpha_1, \dots, \alpha_h$ to define $c_{ij}(n)$ and $b_{ij}(n)$. Let

$\text{Hom}(E_i, E_j)$ be the group of homomorphisms (as abelian varieties over \bar{F}_p) of E_i to E_j .

PROPOSITION 1. *The notation being as above, for each non-negative integer n , we have*

$$\#\{\lambda \in \text{Hom}(E_i, E_j) \mid \text{deg}(\lambda) = n\} = c_{ij}(n)$$

where deg denotes the degree of homomorphisms, and $\#S$ denotes the cardinality of S .

PROOF. This is obvious if $n=0$, and hence we assume that $n>0$. Take a positive integer M so that $a_i^{-1}a_jM = \mathfrak{b}$ is integral. Let $\lambda_{\mathfrak{b}}$ be a \mathfrak{b} -multiplication of (E_i, θ_i) to a \mathfrak{b} -transform (E_j, θ_j) of (E_i, θ_i) . Then by [14] 7.4 Prop. 13, we have $\text{Hom}(E_i, E_j) = \lambda_{\mathfrak{b}} \circ \theta_i(\mathfrak{b}^{-1})$. By [14] 7.2 Prop. 10, we conclude that $\#\{\lambda \in \text{Hom}(E_i, E_j) \mid \text{deg}(\lambda) = n\} = \#\{a \in \mathfrak{b}^{-1} \mid N(\mathfrak{b}a) = n\} = \#\{A \in (a_i^{-1}a_j)^{-1} \mid N(a_i^{-1}a_jA) = n\}$.

Q. E. D.

The following two corollaries follow at once from the above.

COROLLARY 1. *For a positive integer n which is prime to p , $b_{ij}(n)$ is equal to the number of subgroups C of $E_i(\bar{F}_p)$ of order n such that the quotients E_i/C are isomorphic to E_j .*

COROLLARY 2. *As formal power series in q , we have*

$$e_j \mathcal{G}_{ij} = \sum q^{\text{deg}(\lambda)}$$

where the summation in the right hand side is taken over $\text{Hom}(E_i, E_j)$.

As an illustration, let us describe a result of Pizer [9] (Th. 3.2) in terms of elliptic curves. It is known that there always exists a supersingular j -invariant which is contained in F_p . Take one such j -invariant, and call it j_1 . Suppose that there exists a supersingular j -invariant (say) j_2 which is not contained in F_p . Then $j_3 = j_2^2$ is also supersingular. We obviously have a degree preserving bijective map from $\text{Hom}(E_1, E_2)$ to $\text{Hom}(E_1, E_3)$, and hence $\mathcal{G}_{12} = \mathcal{G}_{13}$. This happens if and only if the genus of the curve $\overline{H/\Gamma^*(p)}$ is not zero, where $\Gamma^*(p)$ is the subgroup of $GL_2(\mathbb{Q})$ generated by $\Gamma_0(p)$ and $\begin{bmatrix} 0 & -1 \\ p & 0 \end{bmatrix}$, or equivalently if and only if $p < 37$ or $p = 41, 47, 59$ or 71 .

§ 2. Modular forms mod p .

2.1. In this section, we recall known results for later use. We fix a prime number $p \geq 5$ and put $R = \{a/b \mid a, b \in \mathbb{Z}, (p, b) = 1\}$. For an R -scheme X (resp.

an R -morphism f of R -schemes), we denote by X_s (resp. f_s) the base change of X (resp. f) from R to \mathbf{F}_p . Let $X_0(N)_Q$ be the modular curve over \mathbf{Q} associated to $\Gamma_0(N)$; it is a complete non-singular curve over \mathbf{Q} whose field of rational functions is isomorphic to $\mathbf{Q}(J(z), J(Nz))$, where $J(z) = q^{-1} + 744 + \dots$ is the usual elliptic modular function. We denote by $X_0(N)_R$ or simply by $X_0(N)$ the modular curve over R associated to $\Gamma_0(N)$; $X_0(1)$ is the projective J -line over R , and $X_0(N)$ is the normalization of $X_0(1)$ in $X_0(N)_Q$. For a positive integer n , there is the natural projection $c_N^n: X_0(Nn) \rightarrow X_0(N)$. Also there is an R -morphism $d_N^n: X_0(Nn) \rightarrow X_0(N)$ which, on the general fibre, corresponds to the map of function fields which sends $J(z)$ (resp. $J(Nz)$) to $J(nz)$ (resp. $J(Nnz)$).

Suppose now that N is prime to p . Then $X_0(N)$ is smooth over R (Igusa's theorem). On the other hand, the closed fibre $X_0(pN)_s$ of $X_0(pN)$ has two irreducible components $C_1(pN)$ and $C_2(pN)$ both of which are isomorphic to $X_0(N)_s$. We henceforth identify $C_i(pN)$ with $X_0(N)_s$ so that $c_{N,s}^n$ can be identified with the identity morphism (resp. the Frobenius morphism) on $C_1(pN)$ (resp. $C_2(pN)$), and that $d_{N,s}^n$ can be identified with the Frobenius morphism (resp. the identity morphism) on $C_1(pN)$ (resp. $C_2(pN)$). $C_1(pN)$ and $C_2(pN)$ meet precisely at the mutually \mathbf{F}_p -conjugate "supersingular points" transversally. (For these facts, see Deligne, Rapoport [2] VI, 6 and Ihara [7] §5.)

2.2. Let $M_2(\Gamma_0(p))$ be the space of automorphic forms of weight 2 with respect to $\Gamma_0(p)$. We denote by $M_2(\Gamma_0(p))_R$ the elements of $M_2(\Gamma_0(p))$ whose q -expansions (at infinity) have coefficients in R , and by $\tilde{M}_2(\Gamma_0(p))$ the subspace of $\mathbf{F}_p[[q]]$ obtained by reduction mod p (of coefficients) of the elements of $M_2(\Gamma_0(p))_R$. Thus $\tilde{M}_2(\Gamma_0(p))$ is the space of modular forms mod p of weight 2 with respect to $\Gamma_0(p)$ in the sense of Serre [11] and [12].

It is known that there is the sheaf of regular differentials $\Omega = \Omega_{X_0(M)}$ on $X_0(M)$ provided that M is not divisible by p^2 ([2] I, 2). $\Omega = \Omega_{X_0(M)/R}^1$ if M is not divisible by p . If $M = pN$ with N prime to p , then $H^0(X_0(pN)_s, \Omega_s)$ can be identified with the pairs (ω_1, ω_2) where ω_i are 1-forms on $C_i(pN)$ which are regular except for possible simple poles at supersingular points, and satisfy $\text{Res}_{P_1}(\omega_1) = -\text{Res}_{P_2}(\omega_2)$ if $C_1(pN)$ and $C_2(pN)$ meet at $P_1 \in C_1(pN)$ and $P_2 \in C_2(pN)$. The relation between Ω and modular forms is discussed full in details by Mazur [8] II. We recall that, under the terminology of [8], $M_2(\Gamma_0(p))_R$ is isomorphic to $H^0(X_0(p), \Omega(\text{cusps}))$ ([8] Lemma 4.6). The isomorphism is given by:

$$M_2(\Gamma_0(p))_R \ni f \mapsto f \frac{dq}{q} \in H^0(X_0(p), \Omega(\text{cusps})).$$

From this, we easily obtain the following result of Atkin and Serre.

PROPOSITION 2 (Atkin, Serre; cf. Atkin [1]). *Let j_1, \dots, j_h be the supersingular j -invariants in characteristic p . Then the q -expansions of $f_i = \frac{1}{j_i - \check{J}}$ $\times q \frac{d\check{J}}{dq}$ ($1 \leq i \leq h$) form an \mathbf{F}_{p^2} -basis of $\tilde{M}_2(\Gamma_0(p)) \otimes_{\mathbf{F}_p} \mathbf{F}_{p^2}$, where $\check{J} = q^{-1} + 744 + \dots \in \mathbf{F}_p[[q]]$ is the reduction mod p of J .*

PROOF. Let $C_1(p)$ be the irreducible component of $X_0(p)_s$ which meet the cusp section at infinity. Then the differentials $\frac{d\check{J}}{j_i - \check{J}}$ ($1 \leq i \leq h$) on $C_1(p) \cong X_0(1)_s$ form an \mathbf{F}_{p^2} -basis of the restriction of $H^0(X_0(p)_s, \Omega(\text{cusps})_s) \otimes_{\mathbf{F}_p} \mathbf{F}_{p^2}$ to $C_1(p)$. The assertion follows from this and the above remark. Q. E. D.

REMARK. One can also prove Prop. 2 by using the results of Serre and Swinnerton-Dyer [11], [12]. Indeed, by [12] Th. 11, we have $\tilde{M}_2(\Gamma_0(p)) = \tilde{M}_{p+1}(\Gamma_0(1))$. In view of [11] Th. 1, it is therefore enough to show that Af_i are isobaric polynomials of weights $p+1$ in Q and R , where Q, R and A are as in [11] §1. This can be done by a direct computation using the explicit formula of A (or the Hasse invariant; cf. [11] Th. 3) given by Deuring [3] 8.2.

2.3. Let l be a prime number which is prime to p . Then we have two morphisms $c = c_p^l: X_0(pl) \rightarrow X_0(p)$ and $d = d_p^l: X_0(pl) \rightarrow X_0(p)$. Let c_c and d_c be their base changes by $\text{Spec}(\mathbf{C}) \rightarrow \text{Spec}(\mathbf{R})$. We know that the Hecke operator $T(l)$ on $M_2(\Gamma_0(p)) \cong H^0(X_0(p) \otimes_{\mathbf{R}} \mathbf{C}, \Omega_{X_0(p) \otimes_{\mathbf{R}} \mathbf{C}}(\text{cusps}))$ is given by the formula: $(f|T(l)) \frac{dq}{q} = d_{c_*} \circ c_c^* \left(f \frac{dq}{q} \right)$ for $f \in M_2(\Gamma_0(p))$, where c_c^* is the pullback, and d_{c_*} is the trace (cf. Serre [10] II 12; cf. also below). This can be “descended” to \mathbf{R} as follows. First note that c and d are étale around the maximal points of the closed fibre. Take an open affine subscheme $Y = \text{Spec}(D)$ of $X_0(p)$ which contains the maximal points of $X_0(p)_s$ but which does not contain cusps and supersingular points. Let $X = \text{Spec}(C)$ be the normalization of Y in $X_0(pl)$ via d . Taking Y suitably, we may assume that C is a finite étale D -algebra. On the other hand, the restriction of the sheaf of regular differentials to X (resp. Y) is canonically isomorphic to $\Omega_{X/\mathbf{R}}^1$ (resp. $\Omega_{Y/\mathbf{R}}^1$). Since C is locally free of constant rank over D , we can define the trace from C to D in the usual manner, and hence, tensoring $\Omega_{D/\mathbf{R}}^1$, we obtain the trace: $\Omega_{C/\mathbf{R}}^1 \cong \Omega_{D/\mathbf{R}}^1 \otimes_D C \rightarrow \Omega_{D/\mathbf{R}}^1$. This gives $d_*: H^0(X, \Omega_{X/\mathbf{R}}^1) \rightarrow H^0(Y, \Omega_{Y/\mathbf{R}}^1)$. Combining this map with the pullback c^* of differentials, we obtain an \mathbf{R} -linear endomorphism of $M_2(\Gamma_0(p))_{\mathbf{R}} \cong H^0(X_0(p), \Omega(\text{cusps}))$ because $M_2(\Gamma_0(p))_{\mathbf{R}}$ is stable under Hecke operators. This endomorphism will be also denoted by $T(l)$.

REMARK. Actually, there is the trace morphism: $d_*\Omega_{X_0(p^l)} \rightarrow \Omega_{X_0(p)}$ which extends the above one, defined in a general context (Hartshorne [6] III §8). However, the above elementary (resp. down to earth) description is sufficient (resp. rather convenient) for our purpose.

§3. Theta series mod p .

In this section, we again assume that $p \geq 5$. Let $f = (f_1, \dots, f_h)$ be the vector whose components are $f_i \in \tilde{M}_2(\Gamma_0(p)) \otimes_{F_p} F_{p^2}$. For a positive integer n which is prime to p , we denote by $T(n)$ the Hecke operator acting on $\tilde{M}_2(\Gamma_0(p))$. Put $f|T(n) = (f_1|T(n), \dots, f_h|T(n))$.

PROPOSITION 3. *The notation being as above, let n be a positive integer which is prime to p . Then we have*

$$f|T(n) = \tilde{B}(n)f$$

where $\tilde{B}(n)$ denotes the reduction mod p of the Brandt matrix.

PROOF. Since the Hecke operators and the Brandt matrices satisfy the same recursive relation (Eichler [4] formulas (24) and (25)), it is enough to prove the above assertion when n is a prime number l . As before, let $C_1(p)$ be the irreducible component of $X_0(p)_s$ which meets the cusp section at infinity, and $C_1(pl)$ the irreducible component of $X_0(pl)_s$ above $C_1(p)$. Then c_p^l and d_p^l induce morphisms $\gamma: C_1(pl) \rightarrow C_1(p)$ and $\delta: C_1(pl) \rightarrow C_1(p)$. Let δ_* be the trace for differentials relative to δ . Since the trace considered in §2 commutes with the base change from R to F_p (or F_{p^2}), our assertion will follow from the equality: $\delta_* \circ \gamma^* \left(f_i \frac{dq}{q} \right) = \sum_{k=1}^h \tilde{b}_{ik}(l) f_k \frac{dq}{q}$, which we now propose to prove.

As in §2, we identify $C_1(p)$ (resp. $C_1(pl)$) with $X_0(1)_s$ (resp. $X_0(l)_s$). Then γ (resp. δ) is identified with $c_{1,s}^l$ (resp. $d_{1,s}^l$). Let $\Phi_i(X, \check{J})=0$ be the transformation equation of degree l in characteristic p ; the field generated over $F_p(\check{J})$ by one of its root is isomorphic to the function field of $X_0(l)_s$. From the definition of the trace, we easily see that $\delta_* \circ \gamma^* \left(f_i \frac{dq}{q} \right) = -\frac{d\Phi_i(j_i, \check{J})}{d\check{J}} \frac{d\check{J}}{\Phi_i(j_i, \check{J})} = -d \log \Phi_i(j_i, \check{J})$, which is equal to $-d \log \Phi_i(\check{J}, j_i)$ by the symmetry of Φ_i . But the roots of $\Phi_i(\check{J}, j_i)=0$ with respect to \check{J} consist precisely of the set of j -invariants of the elliptic curves that are obtained by dividing E_i by cyclic subgroups of order l of $E_i(\bar{F}_p)$; i. e. $\Phi_i(\check{J}, j_i) = \prod_{k=1}^h (\check{J} - j_k)^{b_{ik}(l)}$ by Cor. 1 to Prop. 1.

This shows that $-d \log \Phi_i(\check{J}, j_i) = \sum_{k=1}^h \tilde{b}_{ik}(l) f_k \frac{dq}{q}$. Q. E. D.

Let $f = \sum_{n=0}^{\infty} c(n)q^n$ be the q -expansion of f with $c(n) \in (\mathbb{F}_{p^2})^h$. Then we obtain the following

THEOREM. *Suppose that $p \geq 5$, and let $\tilde{\Theta}$ be the reduction mod p of Θ . Then we have $\tilde{\Theta}c(1) = f$. Especially, the theta series mod p $\tilde{\mathcal{G}}_{i,j}(1 \leq i, j \leq h)$ span $\tilde{M}_2(\Gamma_0(p))$.*

PROOF. By Prop. 3 and the well-known relation between the Hecke operators and the coefficients of q -expansions (cf. Shimura [13] 3.5), we have $c(n) = \tilde{B}(n)c(1)$ for all n which are prime to p . Therefore each component of $\tilde{\Theta}c(1) - f$ is a power series in q^p . But it can be considered as a modular form mod p of weight $p+1$ with respect to $\Gamma_0(1)$ ([12] Th. 11). If it were not zero, then its filtration is not divisible by p , and hence applying the operator $\theta = q \frac{d}{dq}$, we get a contradiction by [11] Cor. 3 to Th. 5. This shows that $\tilde{\Theta}c(1) - f = 0$.
 Q. E. D.

REMARK. Explicitly, we have $c(1) = (j_1 - 744, \dots, j_h - 744)$.

References

- [1] Atkin, A.O.L., Modular forms of weight one and supersingular equations, Report for U.S.-Japan Seminar on Applications of Automorphic Forms to Number Theory, Ann Arbor, 1975.
- [2] Deligne, P. et M. Rapoport, Les schémas de modules de courbes elliptiques, Lecture Notes in Math. vol. 349, Springer, 1973, 143-316.
- [3] Deuring, M., Die Typen der Multiplikatorenringe elliptischer Funktionenkörper, Abh. Math. Sem. Univ. Hamburg **14** (1941), 197-272.
- [4] Eichler, M., Zur Zahlentheorie der Quaternionen-Algebren, J. Reine Angew. Math. **195** (1956), 127-151.
- [5] Eichler, M., Über die Darstellbarkeit von Modulformen durch Thetareihen, J. Reine Angew. Math. **195** (1956), 156-171.
- [6] Hartshorne, R., Residues and Duality, Lecture Notes in Math. vol. 20, Springer, 1966.
- [7] Ihara, Y., On modular curves over finite fields, Proc. Internat. Colloq. on Discrete Subgroups of Lie groups, Bombay, Oxford Univ. Press, 1973, 161-202.
- [8] Mazur, B., Modular curves and the Eisenstein ideals, Inst. Hautes Études Sci. Publ. Math. **47** (1977), 33-186.
- [9] Pizer, A., A note on a conjecture of Hecke, Pacific J. Math. **79** (1978), 541-548.
- [10] Serre, J.-P., Groupes algébriques et corps de class, Hermann, 1959.
- [11] Serre, J.-P., Congruences et formes modulaires, Sémin. Bourbaki exp. 416, 1972.
- [12] Serre, J.-P., Formes modulaires et fonctions zêta p -adiques, Lecture Notes in Math. vol. 350, Springer, 1973, 191-268.
- [13] Shimura, G., Introduction to the arithmetic theory of automorphic functions, Publ. Math. Soc. Japan, Iwanami Shoten and Princeton Univ. Press, 1971.

- [14] Shimura, G. and Y. Taniyama, Complex multiplications of abelian varieties and its applications to number theory, Publ. Math. Soc. Japan, no. 6, 1961.

(Received June 17, 1981)

Department of Mathematics
Kyoto University
Kitashirakawa, Kyoto
606, Japan