Modular Forms of Weight One

Pierre DELIGNE and Jean-Pierre SERRE

To Henri Cartan, on the occasion of his 70^{th} birthday

Introduction[1]

The decomposition into an Euler product and the functional equation of the Dirichlet series associated by Hecke to modular forms of weight one suggests that these correspond to Artin L-functions of degree 2 over $\mathbb{Q}$, otherwise known as Galois representations $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{C})$. It is this correspondence, conjectured by Langlands, which we establish here.

The first three sections are preliminary. The fourth section contains the statement of the main theorem and some corollaries. The proof occupies sections 5 through 9. The idea is as follows: one begins by constructing, for each prime number ℓ, a representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ in characteristic ℓ (see §6); we then show that the image of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is "small" under each of these representations, which permits us to lift this to characteristic 0 and obtain the complex representation we seek. The "smallness" in question itself results from an increase of the average of the eigenvalues of the Hecke operators (Rankin, §5). Section 9 contains an estimate of the coefficients of the modular forms of weight one.

We note that we have used an essential point (§6, Thm. 6.1) of some results demonstrated by one of us (P. Deligne), but of which no complete proof has been published; this proof should appear in SGA 5, so we ask that the reader accept this for now.

§1 Modular Forms (Analytic Properties)

1.1 Let $N \geq 1$ be an integer. One associates the subgroups

$$\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N)$$

[1] Translation by Craig Citro and Curtis Paul

of $\mathrm{SL}_2(\mathbb{Z})$ defined by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(N) \iff a \equiv d \equiv 1 \mod N \text{ and } b \equiv c \equiv 0 \mod N,$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_1(N) \iff a \equiv d \equiv 1 \mod N \text{ and } c \equiv 0 \mod N,$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) \iff c \equiv 0 \mod N.$$

1.2 Let f be a function on the upper half-plane $\mathfrak{h}$. If $k \in \mathbb{Z}$, and if $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R})$, we set

$$(f|_k\gamma)(z) = (cz+d)^{-k} f(\gamma z) \quad \text{where } \gamma z = \frac{az+b}{cz+d}.$$

Let $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ be a subgroup containing $\Gamma(N)$. We say that f is *modular of weight k on Γ* if:

(1.2.1) $f|_k\gamma = f$ for all $\gamma \in \Gamma$;

(1.2.2) f is holomorphic on $\mathfrak{h}$;

(1.2.3) f is "pointwise holomorphic," i.e. for all $\sigma \in \mathrm{SL}_2(\mathbb{Z})$, we have that $f|_k\sigma$ has a power series expansion in $e^{2\pi iz/N}$ with nonnegative exponents.

When, in (1.2.3), one replaces "nonnegative exponents" with "positive exponents," we call f a *cusp form*.

1.3 Let f be a modular form of weight k on $\Gamma(N)$. For f to be a modular form on $\Gamma_1(N)$, it is necessary and sufficient that $f(z+1) = f(z)$, or that f has a power series of the form

$$\sum_{n=0}^{\infty} a_n q^n, \quad \text{where } q = e^{2\pi iz}.$$

In the sequel, we shall denote this series by $f_\infty(q)$ or simply f.

1.4 Let f be a weight k modular form on $\Gamma_1(N)$. If $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$, the form $f|_k\gamma$ only depends on the image of d in $\left(\mathbb{Z}/N\mathbb{Z}\right)^\times$; we denote this $f|R_d$. One has $f|R_{-1} = (-1)^k f$.

1.5 Let ε be a Dirichlet character mod N, that is, a homomorphism

$$\varepsilon : \left(\mathbb{Z}/N\mathbb{Z}\right)^\times \to \mathbb{C}^\times.$$

We say that ε is *even* (resp. *odd*) if $\varepsilon(-1) = 1$ (resp. $\varepsilon(-1) = -1$).

Let $k \in \mathbb{Z}$ have the same parity as ε (i.e. $\varepsilon(-1) = (-1)^k$). We say that f is a *modular form of type* (k, ε) *on* $\Gamma_0(N)$ if f is a modular form of weight k on $\Gamma_1(N)$ satisfying

$$f|R_d = \varepsilon(d) f$$

for all $d \in \left(\mathbb{Z}/N\mathbb{Z}\right)^\times$, i.e.

$$f\left(\frac{az+b}{cz+d}\right) = \varepsilon(d)(cz+d)^k f(z) \text{ for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N).$$

(Notice that, if ε and k do not have the same parity, this formula would imply $f = 0$.)

Every modular form of weight k on $\Gamma_1(N)$ is a linear combination of forms of type (k, ε_i) on $\Gamma_0(N)$, where the ε_i are the distinct characters on $\left(\mathbb{Z}/N\mathbb{Z}\right)^\times$ of the same parity as k.

This is shown (*cf.* [14], p. IV-13) by noticing that the function

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto d \mod N$$

gives, by passing to the quotient, an isomorphism of

$$\Gamma_0(N)/\Gamma_1(N) \cong \left(\mathbb{Z}/N\mathbb{Z}\right)^\times.$$

1.6 Hecke Operators Let $f = \sum a_n q^n$ be a modular form of type (k, ε) on $\Gamma_0(N)$ and let p prime. We define

$$f|T_p = \sum a_{pn} q^n + \varepsilon(p) p^{k-1} \sum a_n q^{pn} \quad \text{if } p \nmid N, \tag{1.6.1}$$

$$f|U_p = \sum a_{pn} q^n \quad \text{if } p \mid N. \tag{1.6.2}$$

We then obtain another modular form of type (k, ε) on $\Gamma_0(N)$ which is cuspidal if f is.

1.7 Newforms One has the definition of a *newform* (or *primitive form*) of type (k, ε) on $\Gamma_0(N)$ (see [2] for $\varepsilon = 1$, and [5], [12], [13] for the general case).

If $f = \sum_{n=1}^{\infty} a_n q^n$ is such a form, we have $a_1 = 1$, and f is an eigenfunction of the Hecke operators T_p and U_p, with corresponding eigenvalues a_p. This gives a Dirichlet series

$$\Phi_f(s) = \sum a_n n^{-s} \tag{1.7.1}$$

which can be represented as the Euler product

$$\Phi_f(s) = \prod_{p \mid N} \frac{1}{1 - a_p p^{-s}} \prod_{p \nmid N} \frac{1}{1 - a_p p^{-s} + \varepsilon(p) p^{k-1-2s}}. \tag{1.7.2}$$

1.8 If f is as above, and if $p \mid N$, we have that $|a_p|$ is given by [12], [16])

- $|a_p| = 0$ if $p^2 \mid N$ and ε can be defined mod N/p,
- $|a_p| = p^{(k-1)/2}$ if ε cannot be defined mod N/p,
- $|a_p| = p^{k/2-1}$ if $p^2 \nmid N$ and ε can be defined mod N/p.

(It will follow from Thm 4.6 below that the last case cannot occur for $k = 1$.)

1.9 Every form f of type (k, ε) on $\Gamma_0(N)$ can be written

$$f(z) = E(z) + \sum \lambda_i f_i(d_i z),$$

where E is an Eisenstein series, and f_i is a cuspidal newform of type (k, ε) on $\Gamma_0(N_i)$, with $N_i \mid N$ such that ε can be defined mod N_i, and $d_i | N/N_i$. Moreover, this decomposition is unique (in the obvious sense).

§2 Modular Forms (Geometric Properties)

2.1 Let $k, N \in \mathbb{Z}$ be positive, and μ_N the group scheme of N^{th} roots of unity. From the geometric viewpoint, a weight k modular form for $\Gamma_1(N)$ is a rule assigning, for each elliptic curve E with an injection $\alpha : \mu_N \to E$, a section of the k-fold tensor $\omega_E^{\otimes k}$ of ω_E, where ω_E is the dual of the Lie algebra of E.

Remarks:

(a) An *elliptic curve* on a scheme S is a smooth proper map $E \to S$, with a distinguished section $e : S \to E$ of geometric fibres of elliptic curves. When $S = \operatorname{Spec} A$ for a commutative ring A, one says E is an elliptic curve over A. We call $\omega_E = e^* \Omega^1_{E/S}$; for $S = \operatorname{Spec} A$, ω_E is identified with an invertible A-module.

(b) Let R be a commutative ring with N invertible in R. A modular form of weight k on $\Gamma_1(N)$, meromorphic at infinity, defined on R, is a rule associating to each elliptic curve on an R-algebra A an injection $\alpha : \mu_N \to E$, with $f(E, \alpha)$ in $\omega_E^{\otimes k}$. We require that this law be compatible with isomorphisms and extension of scalars.

(c) One says that f is *holomorphic at infinity* if f can be extended to a rule $\tilde{f}$ defined for the pairs (E, α) where E is a generalized elliptic curve ([7], II.1.12) and an injection $\alpha : \mu_N \to E$ whose image meets each irreducible component of each geometric fiber ([7], IV.4.14). If such $\tilde{f}$ exists, it is unique.

Let R be a field. We say f is *cuspidal* if it is holomorphic at infinity and if $\tilde{f}(E, \alpha) = 0$ each time E is a degenerate elliptic curve (i.e. not smooth) on an algebraically closed extension of R.

These notions can also be defined in terms of Laurent series in q ([7], VII, §3)

2.2 Let f be as above. If $d \in \left(\mathbb{Z}/N\mathbb{Z}\right)^\times$, we define the modular form $f|R_d$ by

$$(f|R_d)(E,\alpha) = f(E, d\alpha). \tag{2.2.1}$$

If ε is a homomorphism of $\left(\mathbb{Z}/N\mathbb{Z}\right)^\times$ on $R^\times$, one says that f is *of type* (k,ε) *on* $\Gamma_0(N)$ if $f|R_d = \varepsilon(d)f$ for all $d \in \left(\mathbb{Z}/N\mathbb{Z}\right)^\times$.

2.3 Let $p \nmid N$ be a prime. The Hecke operator T_p is then defined on the spaces of modular forms. If f is such a form, and if (E,α) is defined over an algebraically closed field of characteristic $\neq p$, and we have

$$(f|T_p)(E,\alpha) = \frac{1}{p}\sum_{\varphi} \varphi^*(f(\varphi E, \varphi\circ\alpha)),$$

where φ runs over all isogeny classes of degree p over E (with two isogenies being in the same class if their kernels are the same).

The T_p commute with one another, and with the R_d.

2.4 Let $R = \mathbb{C}$. The map $\alpha : \mu_N \to E$ is then determined by

$$\alpha(e^{2\pi i/N})$$

which must be of order N. To a modular form f as above, one associates a function (also called f) on $\mathfrak{h}$ via

$$f(z) = \frac{f(E_z, 1/N)}{(2\pi i du)^{\otimes k}}, \tag{2.4.1}$$

where E_z is the elliptic curve $\mathbb{C}/(\mathbb{Z}\oplus z\mathbb{Z})$.

Define $f(z) = f_\infty(e^{2\pi i z})$. Then 2.4.1 can be rewritten as:

$$f_\infty(q) = \frac{f(\mathbb{C}^\times/q^{\mathbb{Z}}, \mathrm{Id})}{(dt/t)^{\otimes k}} \quad (0<|q|<1), \tag{2.4.2}$$

where Id denotes the inclusion $\mu_N \hookrightarrow \mathbb{C}^\times$.

This construction identifies the spaces of modular forms in the sense of §2.1, §2.2 with the same spaces of §1; the same follows for the operators T_p and R_d.

2.5 For the definition of the *Tate curve* $\mathbb{G}_m/q^{\mathbb{Z}}$ on the ring $\mathbb{Z}((q)) = \mathbb{Z}[[q]](q^{-1})$, we refer to [7], VII, §1. This curve is equipped with an invariant differential form dt/t, and a natural embedding $\mathrm{Id} : \mu_N \to \mathbb{G}_m/q^{\mathbb{Z}}$. If f is a modular form of weight k on $\Gamma_1(N)$, meromorphic at infinity, and defined on a ring R, we define

$$f_\infty(q) = \frac{f(\mathbb{G}_m/q^{\mathbb{Z}}, \mathrm{Id})}{(dt/t)^{\otimes k}} \in \mathbb{Z}((q))\otimes R \subset R((q)).$$

(In this formula, $\mathbb{G}_m/q^{\mathbb{Z}}$ denotes the curve on $Z((q)) \otimes R$ determined by the Tate curve via extension of scalars.)

Define

$$f_\infty(q) = \sum a_n q^n$$

and

$$(f|R_d)_\infty(q) = \sum a_n(d) q^n, \quad d \in \left(\mathbb{Z}/N\mathbb{Z}\right)^\times;$$

if $p \nmid N$ is prime, one has

$$(f|T_p)_\infty(q) = \sum a_{pn} q^n + p^{k-1} \sum a_n(p) q^{np}. \tag{2.5.1}$$

In particular, if f is of type (k, ε) on $\Gamma_0(N)$, one has

$$(f|T_p)_\infty(q) = \sum a_{pn} q^n + \varepsilon(p) p^{k-1} \sum a_n q^{pn}. \tag{2.5.2}$$

When $R = \mathbb{C}$, $f_\infty(q)$ is the power series in 2.4.2; this is proven in [7], VII, §4 (at least for f holomorphic, the only case which concerns us). The formula 2.5.2 again gives 1.6.1.

2.6 If K is a field of characteristic 0, we denote by S_K the vector space of cusp forms of weight k on $\Gamma_1(N)$ which are defined over K. One has

$$S_K = K \otimes_{\mathbb{Q}} S_{\mathbb{Q}}; \tag{2.6.1}$$

which can be seen by interpreting S_K as the space of sections of an invertible sheaf over the algebraic stack corresponding to $\Gamma_1(N)$. ([7], VII, 3.2).

If K' is a subfield of K, a form $f \in S_K$ belongs to $S_{K'}$ iff the coefficients of the power series $f_\infty(q)$ belong to K'. This can be seen by going back to the case where K is algebraically closed, and noticing that for all automorphisms σ of K over K', the forms f and $\sigma(f)$ have power series which coincide.

Proposition 2.7. *Let L be the set of $f \in S_{\mathbb{C}}$ such that $(f|R_d)_\infty(q) \in Z[[q]]$ for all $d \in \left(\mathbb{Z}/N\mathbb{Z}\right)^\times$. Then:*

(2.7.1) *L is a free $\mathbb{Z}$-module of finite type, stable under the operators T_p and R_d.*

(2.7.2) *For each field K of characteristic 0, we have $S_K = K \otimes L$.*

(2.7.3) *The eigenvalues of the T_p on $S_{\mathbb{C}}$ are integers in a finite extension of $\mathbb{Q}$.*

(2.7.4) *If $f \in S_{\mathbb{C}} = \mathbb{C} \otimes L$ is such that $f|T_p = a_p f$, then, for all automorphisms σ of $\mathbb{C}$, the form $\sigma(f)$ is such that $\sigma(f)|T_p = \sigma(a_p)\sigma(f)$. If f is of type (k, ε) over $\Gamma_0(N)$, then $\sigma(f)$ is of type $(k, \sigma(\varepsilon))$ on $\Gamma_0(N)$.*

If $f \in S_{\mathbb{Q}}$, we have $f|R_d \in S_{\mathbb{Q}}$ for all $d \in \left(\mathbb{Z}/N\mathbb{Z}\right)^\times$, and the series $(f|R_d)_\infty(q)$ belongs to $\mathbb{Z}[[q]] \otimes \mathbb{Q}$, which has bounded denominators. It follows that a nonzero multiple of f belongs to L, where $\mathbb{Q} \otimes L = S_{\mathbb{Q}}$, and by

2.6.1, $K \otimes L = S_K$ for every field K of characteristic 0. That L is of finite type is shown by the fact that the linear forms "n-th coefficients of $(f|R_d)_\infty(q)$" separate the elements of $S_{\mathbb{Q}}$.

The fact that L is stable under R_d (resp. T_p) is evident (resp. results from 2.5.1). The assertions (2.7.3) and (2.7.4) follow.

2.8 The fact that the series $f_\infty(q)$, $f \in S_{\mathbb{Q}}$, is of bounded denominator has here been deduced from the fact that the Tate curve is defined over $\mathbb{Z}((q)) \otimes \mathbb{Q}$. One would have equally been able to use Thm. 3.5.2 of [24], valid when $k \geq 2$, and changing the weight 1 to weight 13 via multiplication by Δ.

§3 Galois Representations

3.1 Let $\overline{\mathbb{Q}}$ be an algebraic closure of $\mathbb{Q}$, and let $G = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. We consider linear representations of G, otherwise known as continuous homomorphisms

$$\rho : G \to \mathrm{GL}_n(k)$$

where k is of one of the following types:

(a) the field $\mathbb{C}$ (with the discrete topology);

(b) a finite field (with the discrete topology);

(c) a finite extension of an ℓ-adic field $\mathbb{Q}_\ell$ (with its natural topology).

In the first two cases, the image of ρ is *finite*.

If p is a prime, one says that ρ is *unramified at* p if it is trivial over the inertial group of a prime of $\overline{\mathbb{Q}}$ lifting p. One then calls $F_{\rho,p}$ the image under ρ of the Frobenius[2] relative to p; it is an element of $\mathrm{GL}_n(k)$, defined up to conjugation. We define

$$\begin{aligned} P_{\rho,p}(T) &= \det(1 - F_{\rho,p}T) \\ &= 1 - \mathrm{Tr}(F_{\rho,p})T + \cdots + (-1)^n \det(F_{\rho,p})T^n. \end{aligned} \tag{3.1.1}$$

Knowing the polynomials $P_{\rho,p}(T)$ almost allows us to reconstruct ρ. More precisely:

Lemma 3.2. *Let X be a set of primes of density 1 and let ρ and ρ' be two semisimple linear representations of G. Suppose that, for all $p \in X$, ρ and ρ' are not ramified, and that the two polynomials $P_{\rho,p}(T) = P_{\rho',p}(T)$ (resp. that $\mathrm{Tr}(F_{\rho,p}) = \mathrm{Tr}(F_{\rho',p})$ when k is of characteristic 0). Then ρ and ρ' are isomorphic.*

This results from the Chebotarev Density Theorem, combined with the fact that a semisimple linear representation of a group is determined up to isomorphism by the corresponding characteristic polynomials (resp. the traces if the characteristic of the field is 0); see [3], §30.16.

Remarks:

[2] We adopt here the conventions of Artin [1]. Our "substitution of Frobenius" is then the element called φ in [6]; its inverse is the "geometric Frobenius."

3.3 In the following, one applies Lemma 3.2 in the particular case when X is the set of prime numbers which do not divide a given integer N; one then says that ρ and ρ' are *not ramified outside of* N.

3.4 When $k = \mathbb{C}$, the condition of semisimplicity is automatically satisfied. Then $\rho(G)$ and $\rho'(G)$ are finite.

§4 Results

(a) Statement of the Principal Theorem

Theorem 4.1. *Let $N \geq 1$ be an integer, ε a Dirichlet character mod N such that $\varepsilon(-1) = -1$, and f a modular form of type $(1, \varepsilon)$ on $\Gamma_0(N)$, not identically zero. Suppose that f is an eigenfunction of T_p, $p \nmid N$, with eigenvalues a_p. Then there exists a linear representation (with $G = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$)*

$$\rho : G \to \mathrm{GL}_2(\mathbb{C})$$

which is unramified outside N such that

$$\mathrm{Tr}(F_{\rho,p}) = a_p \quad \text{and} \quad \det(F_{\rho,p}) = \varepsilon(p) \text{ for all } p \nmid N. \tag{4.1.1}$$

This representation is irreducible iff f is cuspidal.

The proof will be given in §8.

Corollary 4.2. *The eigenvalues a_p are sums of two roots of unity; in particular we have $|a_p| \leq 2$.*

In other words, the Ramanujan-Petersson Conjecture is true for weight one; however, it is known that it is equally true for weights ≥ 2, see [6], 8.2.

Remarks:

4.3 As a result of Lemma 3.2, the representation ρ associated to f by 4.1 is unique, up to isomorphism.

4.4 The formula $\det(F_{\rho,p}) = \varepsilon(p)$ shows that we have

$$\det(\rho) = \varepsilon,$$

by identifying ε with a character $G \to \mathbb{C}^\times$ which corresponds to it by class field theory (it is simply the composition of ε and the homomorphism $G \to \left(\mathbb{Z}/N\mathbb{Z}\right)^\times$ given by the action of G on the N^{th} roots of unity).

4.5 Let c be the element of G corresponding to complex conjugation (for an embedding $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$). Since ε is odd, 4.4 shows that $\det(\rho(c)) = -1$; since c has order 2, this shows that $\rho(c)$ is conjugate to the matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

(b) The Artin Conductor and Local Factors – We retain the hypotheses and notation of Thm. 4.1.

Theorem 4.6. *Suppose that f is a cuspidal newform with coefficients a_n, $n \geq 1$. Let ρ be a corresponding representation of G. Then:*

a. The Artin Conductor of ρ is N;

b. The Artin L*-function* $\mathrm{L}(s, \rho)$ *is equal to*

$$\Phi_f(s) = \sum_{n=1}^{\infty} a_n n^{-s}.$$

(For the definition of the L-series, and the conductor, see [1]).

Corollary 4.7. *The representation ρ is ramified at all the prime divisors of N.*

This follows from (a).

Corollary 4.8. *The function* $\mathrm{L}(s, \rho)$ *is entire.*

[In other words, the "Artin conjecture" is true for ρ; see (c) below.]
In fact, Hecke theory shows that $\Phi_f(s)$ is entire.

Proof of 4.6: This uses the *functional equations* satisfied by $\Phi_f(s)$ and $\mathrm{L}(s, \rho)$ (compare to [9], p. 172-177).

(i) Let $\tilde{f} = \sum \overline{a}_n q^n$. Since f is a newform, there exists $\lambda \neq 0$ such that $f\left(\dfrac{-1}{Nz}\right) = \lambda z \tilde{f}(z)$ (see [12], [13]). By the Mellin transform, we have

$$\Psi_f(1-s) = \mu \tilde{\Psi}_f(s),$$

where

$$\mu = \frac{i\lambda}{\sqrt{N}}$$
$$\Psi_f(s) = N^{s/2}(2\pi)^{-s}\Gamma(s)\Phi_f(s)$$
$$\tilde{\Psi}_f(s) = \Psi_{\tilde{f}}(s).$$

(ii) By 4.5, the "factor at infinity" of $\mathrm{L}(s, \rho)$ is equal to $(2\pi)^{-s}\Gamma(s)$. If M is the conductor of ρ, and if we let

$$\xi(s, \rho) = M^{s/2}(2\pi)^{-s}\Gamma(s)\,\mathrm{L}(s, \rho),$$

then we have

$$\xi(1-s, \rho) = \nu \cdot \xi(s, \overline{\rho})$$

where $\nu \in \mathbb{C}^{\times}$.

(iii) Let

$$F(s) = \frac{(N/M)^{s/2}\Psi_f(s)}{\xi(s,\rho)} \quad \text{and} \quad \tilde{F}(s) = \frac{(N/M)^{s/2}\tilde{\Psi}_f(s)}{\xi(s,\overline{\rho})}.$$

The above formulas show that

$$F(1-s) = \omega \cdot \tilde{F}(s)$$

where $\omega = \mu/\nu$. But, if $p \not| N$ is prime, the p-factors of $\Psi_f(s)$ and of $\xi(s,\rho)$ coincide by 4.1.We then have

$$F(s) = A^s \prod_{p|N} F_p(s),$$

with $A = \sqrt{N/M}$ and

$$F_p(s) = \frac{1 - a_p p^{-s}}{(1 - b_p p^{-s})(1 - c_p p^{-s})},$$

where $1 - a_p p^{-s}$ is the p-factor of $\Psi_f(s)$, and $(1 - b_p p^{-s})(1 - c_p p^{-s})$ that of $\xi(s,\rho)$ (notice that b_p and c_p can be zero). Everything reduces to showing that A and the F_p are equal to 1. For this we use the following elementary lemma:

Lemma 4.9. *Let $G(s) = A^s \prod_p G_p(s)$, $H(s) = A^s \prod_p H_p(s)$ be two finite Euler products. Suppose that*

(4.9.1) $G(1-s) = \omega \cdot H(s)$ *with* $\omega \in \mathbb{C}^\times$;

(4.9.2) *Each of the G_p, H_p is a finite product of terms of the form*

$$(1 - \alpha_p^{(i)} p^{-s})^{\pm 1}, \text{ with } |\alpha_p^{(i)}| < \sqrt{p}.$$

One then has $A = 1$, and $G_p = H_p = 1$ for all p.

If H_p is not equal to 1, the function H has an infinite number of zeros (or poles) of the form

$$\frac{\log(\alpha_p^{(i)}) + 2\pi i n}{\log p}, \qquad n \in \mathbb{Z},$$

and we easily see that these cannot be all of the zeros (or poles) of $G(1-s)$; the hypothesis $|\alpha_p^{(i)}| < \sqrt{p}$ assures, in fact, that each of the $\alpha_p^{(i)}$ cannot be equal to a $p/\alpha_p^{(j)}$.

(iv) It still remains to show that a_p, b_p, c_p, and their conjugates, satisfy (4.9.2), i.e. are less than $\sqrt{p}$ in absolute value. It is clear for b_p and c_p, which are either 0 or roots of unity. For a_p, one can invoke 1.8, which shows that $|a_p| \leq 1$; one can also, if one prefers, use the Rankin inequality:

$$|a_n| = O(n^{1/2-1/5}) \qquad \text{(see [18])};$$

by using $n = p^m$, and by noting that $a_n = (a_p)^m$, we have that

$$|a_p| \leq p^{1/2-1/5} < p^{1/2}.$$

This finishes the proof of 4.6.

(c) Characterization of Representations attached to forms of weight one – We keep the notations of (a), and suppose that f is cuspidal. The corresponding representation $\rho : G \to \mathrm{GL}_2(\mathbb{C})$ then has the following properties:

(i) ρ is irreducible (4.1);

(ii) $\det \rho$ is an odd character (4.4);

(iii) For every continuous character $\chi : G \to \mathbb{C}^\times$, the Artin L-function $\mathrm{L}(s, \rho \otimes \chi)$ is an entire function – this follows from 4.8, applied to the cusp form

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

Reciprocally:

Theorem 4.10. (Weil-Langlands): *Let $\rho : G \to \mathrm{GL}_2(\mathbb{C})$ be a continuous representation of $G = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ satisfying conditions (i), (ii), and (iii) above. Let $\varepsilon = \det \rho$, $N =$ conduct.ρ,*

$$L(s, \rho) = \sum a_n n^{-s}, \quad f = \sum a_n q^n.$$

Then f is a cuspidal newform of type $(1, \varepsilon)$ on $\Gamma_0(N)$, and ρ is the representation attached to f.

After Langlands (see [27], pgs. 152, 160), the constants of the functional equations of the series $\sum a_n \chi(n) n^{-s}$ verify the necessary identity so that one can apply the characterization of the modular forms due to Hecke-Weil ([12], [26]). It follows that f is modular of type $(1, \varepsilon)$ on $\Gamma_0(N)$; it is clear that f is an eigenfunction of the T_p and of the U_p, and that the representation associated to it is isomorphic to ρ. After 4.1, f is cuspidal. Let f' be the unique cuspidal newform (on $\Gamma_0(N')$, where N' is a convenient divisor of N) such that $f'|T_p = a_p f'$ for $p \not| N$. As seen in 4.6, the Dirichlet series associated to f' is $\mathrm{L}(s, \rho) = \sum a_n n^{-s}$. It follows that $f' = f$, which shows that f is a newform.

Remarks:

4.11 One will find in [27], p.163, a generalization of Thm. 4.10 to global fields.

4.12 Condition (iii) (the Artin Conjecture for $\rho \otimes \chi$) can be replaced by the following weaker condition:

(iii'): There exists an integer $M \geq 1$ such that, for every character χ of conductor prime to M, the function $\mathrm{L}(s, \rho \otimes \chi)$ is entire.

This is a result of [26] (see also [12]).

4.13 If the Artin Conjecture is true, the theorems above give a bijection between "classes of irreducible representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ of degree 2 with odd determinant" and "cuspidal newforms of weight 1."

4.14 Thm. 4.6 even gives a way to *verify* the Artin Conjecture in some cases. If one is given a representation ρ satisfying (i) and (ii), of conductor N and determinant ε, one can numerically determine the coefficients a_n of the series $\mathrm{L}(s,\rho) = \sum a_n n^{-s}$ for n smaller than a given integer A, and one can look to construct a cuspidal newform f of type $(1,\varepsilon)$ on $\Gamma_0(N)$ for which the series begins $\sum_{n\le A} a_n q^n$. If A is large enough, for example

$$A \ge \frac{N}{12} \prod_{p\,|N} \left(1 + \frac{1}{p}\right),$$

such a form is unique, if it exists (if it does not exist, the Artin Conjecture is false). Once we obtain f, it corresponds to a representation ρ_f; if one can show ρ_f is isomorphic to ρ, it must be that ρ satisfies (iii).

Examples: If ρ is as above, the image of ρ in the group

$$\mathrm{PGL}_2(\mathbb{C}) = {}^{\mathrm{GL}_2(\mathbb{C})}\!/_{\mathbb{C}^\times}$$

is either a dihedral group, or one of the groups $\mathfrak{A}_4$, $\mathfrak{A}_5$, or $\mathfrak{S}_4$ ([22], Prop. 16). In the *dihedral* case, ρ is induced by a representation of degree 1 of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\sqrt{d}))$, where $\mathbb{Q}(\sqrt{d})$ is a quadratic extension of $\mathbb{Q}$. Condition (iii) is then verified, and ρ must correspond to a cusp form; this is a linear combination of *theta series* for binary forms of discriminant d, see [9], p. 428-460. Some *non-dihedral* examples have recently been constructed by Tate (for $N = 133, 229, 283, 331, \dots$).

For one of these examples (that of $N = 133$, which corresponds to a group $\mathfrak{A}_4$), Tate, with the help of Atkin, *et al.*, has been able to extend the method sketched in 4.14, and prove the existence of a corresponding modular form – and also the Artin Conjecture for the representation in question.

§5 Use of a result of Rankin

Proposition 5.1. *Let f be a cusp form of type (k,ε) for $\Gamma_0(N)$, not identically zero. Suppose f is an eigenfunction of the T_p, $p \nmid N$, with eigenvalues a_p. Then the series*

$$\sum_{p\nmid N} |a_p|^2 p^{-s}$$

converges for $\mathrm{Re}\, s > k$, *and*

$$\sum_{p\nmid N} |a_p|^2 p^{-s} \le \log\left(\frac{1}{s-k}\right) + O(1) \tag{5.1.1}$$

for $s \to k$.

5.2 (Proof of 5.1) We immediately reduce to the case where f is a newform $\sum_{n=1}^{\infty} a_n q^n$. For each $p \not| N$, let $\varphi_p \in \mathrm{GL}_2(\mathbb{C})$ with $\mathrm{Tr}(\varphi_p) = a_p$ and $\det(\varphi_p) = \varepsilon(p)p^{k-1}$. The Dirichlet series

$$\Phi_f(s) = \sum_{n=1}^{\infty} a_n n^{-s}$$

can then be written

$$\Phi_f(s) = \prod_{p\,|N} (1 - a_p p^{-s})^{-1} \prod_{p \not| N} \det(1 - \varphi_p p^{-s})^{-1},$$

cf. (1.7.2). Let

$$\mathrm{L}(s) = \prod_{p \not| N} \det(1 - \varphi_p \otimes \overline{\varphi}_p p^{-s})^{-1}.$$

This is an Euler product with four factors: if we call λ_p, μ_p the eigenvalues of φ_p, we have

$$\mathrm{L}(s) = \prod_{p \not| N} [(1 - \lambda_p \overline{\lambda}_p p^{-s})(1 - \lambda_p \overline{\mu}_p p^{-s})(1 - \mu_p \overline{\lambda}_p p^{-s})(1 - \mu_p \overline{\mu}_p p^{-s})]^{-1}.$$

Using the formula

$$\lambda_p \overline{\lambda}_p \mu_p \overline{\mu}_p = |\varepsilon(p)p^{k-1}|^2 = p^{2k-2},$$

one shows (see, for example, [10], p. 33, or [12], [15]) that

$$\mathrm{L}(s) = H(s)\zeta(2s - 2k + 2) \left(\sum_{n=1}^{\infty} |a_n|^2 n^{-s} \right),$$

with

$$H(s) = \prod_{p\,|N} (1 - p^{-2s+2k-2})(1 - |a_p|^2 p^{-s}).$$

As in [18] (*cf.* also [12], [13], and [15]), the series $\sum |a_n|^2 n^{-s}$ converges for $\mathrm{Re}\, s > k$ and its product by $\zeta(2s - 2k + 2)$ extends to a meromorphic function on the whole complex plane, with a unique pole at $s = k$. Since $|a_p| < p^{k/2}$ if $p\,|\,N$ (*cf.* 1.8), the function $H(s)$ is holomorphic and nonzero for $\mathrm{Re}\, s \geq k$. This results in $\mathrm{L}(s)$ being meromorphic in the whole complex plane and holomorphic for $\mathrm{Re}\, s \geq k$, with the sole exception of $s = k$, which is a simple pole; moreover, we have $\mathrm{L}(s) \neq 0$ for $\mathrm{Re}\, s > k$, since this holds for $H(s)$, $\zeta(2s - 2k + 2)$, and $\sum |a_n|^2 n^{-s}$.

Define

$$g_m(s) = \sum_{p \not| N} |\,\mathrm{Tr}(\varphi_p^m)|^2 \frac{p^{-ms}}{m} \quad \text{and} \quad g(s) = \sum_{m=1}^{\infty} g_m(s).$$

The series $g(s)$ is a Dirichlet series with coefficients ≥ 0. For s sufficiently large, a quick calculation shows that it is equal to $\log \mathrm{L}(s)$. Since $\mathrm{L}(s)$ is holomorphic and nonvanishing for $\mathrm{Re}\, s > k$, we have by a classic lemma of Landau ([21], p. 112) that $g(s)$ converges for $\mathrm{Re}\, s > k$. From the fact that $\mathrm{L}(s)$ has a simple pole at $s = k$, we have

$$g(s) = \log\left(\frac{1}{s-k}\right) + O(1) \quad \text{as } s \to k.$$

But $g_1(s) = \sum |a_p|^2 p^{-s}$ is clearly $\leq g(s)$. Hence

$$\sum |a_p|^2 p^{-s} = \log\left(\frac{1}{s-k}\right) + O(1) \quad \text{as } s \to k.$$

5.3 Remarks – One can strengthen Prop. 5.1 in different ways. First, assuming the Petersson Conjecture, an easy overestimation shows that the series

$$\sum_{p \nmid N} \sum_{m \geq 2} |\operatorname{Tr}(\varphi_p^m)|^2 \frac{p^{-ms}}{m} = g_2(s) + g_3(s) + \cdots$$

converges for $\mathrm{Re}\, s \geq k$, and this allows us to replace the inequality 5.1.1 with the equality

$$\sum |a_p|^2 p^{-s} = \log\left(\frac{1}{s-k}\right) + O(1) \quad \text{as } s \to k. \tag{5.3.1}$$

On the other hand, an argument of Hadamard-de la Vallée Poussin shows that $\mathrm{L}(s) \neq 0$ for all s such that $\mathrm{Re}\, s \geq k$ (including the critical line $\mathrm{Re}\, s = k$), and an application of the theorem of Wiener-Ikehara on $\mathrm{L}'(s)/\mathrm{L}(s)$, one has

$$\sum_{p \leq x} |a_p|^2 p^{-(k-1)} \sim x \quad \text{for } x \to \infty, \tag{5.3.2}$$

cf. Rankin [19].

5.4 Application to Forms of Weight 1 Let P be the set of prime numbers, $X \subseteq P$. Define

$$\mathrm{dens}\,.\sup X = \limsup_{s \to 1, s > 1} \left(\frac{\sum_{p \in X} p^{-s}}{\log(1/(s-1))}\right). \tag{5.4.1}$$

This is the *upper density* of X; it lies between 0 and 1.

Proposition 5.5. *If we keep the hypotheses of 5.1, and suppose moreover that the weight k of f is 1, then for all $\eta > 0$, there exists a set X_η of prime numbers and a finite subset Y_η of $\mathbb{C}$ such that*

$$\mathrm{dens}\,.\sup X_\eta \leq \eta \quad \textit{and} \quad a_p \in Y_\eta \textit{ for all } p \notin X_\eta.$$

After 2.7, the a_p are integers in a finite extension K of $\mathbb{Q}$. If $c \geq 0$ is a constant, we define

$$Y(c) = \{a \in \mathfrak{O}_K \mid |\sigma(a)|^2 \leq c \quad \forall \sigma : K \to \mathbb{C} \text{ embedding}\};$$

this is a finite set. We call $X(c)$ the set of p such that $a_p \notin Y(c)$; it suffices to show that $\operatorname{dens}.\sup X(c) \leq \eta$ for c sufficiently large.

Yet we know (2.7) that the $\sigma(a_p)$ are also eigenvalues of the T_p of weight one. We have seen (5.1.1) that we then have

$$\sum_\sigma \sum_p |\sigma(a_p)|^2 p^{-s} \leq r \log\left(\frac{1}{s-1}\right) + O(1) \quad \text{as } s \to 1,$$

where $r = [K : \mathbb{Q}]$. Since $\sum_\sigma |\sigma(a_p)|^2 \geq c$ if $p \in X(c)$, we conclude that

$$c \sum_{p \in X(c)} p^{-s} \leq r \log\left(\frac{1}{s-1}\right) + O(1) \quad \text{as } s \to 1,$$

where

$$\operatorname{dens}.\sup X(c) \leq \frac{r}{c},$$

and it then suffices to take $c \geq r/\eta$.

5.6 Remarks – Using (5.3.2) in place of (5.1.1) in the above proof, one can replace the "analytic" density 5.4.1 with the "natural" density (*cf.* [21], VI, n° 4.5). In any case, 5.5 has only temporary interest: once Thm. 4.1 has been proven, we'll know that the set of a_p is *finite*.

§6 ℓ-adic Representations and Reduction $\bmod \ell$

(a) ℓ-adic Representations: We use the following result:

Theorem 6.1. *Let f be a modular form of type (k, ε) on $\Gamma_0(N)$, not identically zero. Let $k \geq 2$ and say f is an eigenfunction of the T_p, $p \nmid N$, with eigenvalues a_p. Let K be a number field containing the a_p and the $\varepsilon(p)$,* cf. *(2.7.3). Let λ be a finite prime of K, with residual characteristic ℓ, and let K_λ the completion of K in λ. Then there exists a continuous semisimple linear representation*

$$\rho_\lambda : G \to \mathrm{GL}_2(K_\lambda),$$

where $G = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, which is unramified outside $N\ell$, and such that

$$\operatorname{Tr}(F_{\rho_\lambda,p}) = a_p \quad \text{and} \quad \det(F_{\rho_\lambda,p}) = \varepsilon(p)p^{k-1} \text{ for } p \nmid N\ell. \tag{6.1.1}$$

By 3.2, the condition (6.1.1) determines ρ_λ uniquely up to isomorphism.

6.2 If f is an Eisenstein series, the above can be deduced immediately from the work of Hecke ([9], p. 690) by taking ρ_λ to be the direct sum of two degree 1 representations. When f is cuspidal, 6.1 is shown in one particular case in [4]. The general case is not much more difficult. It is shown by another method (inspired by Ihara) and in another language, by Langlands [11] (where, however, one must accept a "trace formula" which seems reasonable, but is not proven). In a forthcoming work by one of us, this result will be proven again by a method due to Piateckii-Shapiro [17].

Corollary 6.3. *Let $(f, N, k, \varepsilon, (a_p))$ and $(f', N', k', \varepsilon', (a'_p))$ be as in Thm. 6.1. If the set of prime numbers where $a_p = a'_p$ has density 1, then $k = k'$, $\varepsilon = \varepsilon'$, and $a_p = a'_p$ for all $p \nmid NN'$.*

In fact, the representations attached to f and f' (for the same choice of K and λ) are isomorphic by 3.2.

Remarks

6.4 The image of G under ρ_λ is a compact subgroup of $\mathrm{GL}_2(K_\lambda)$, hence an *ℓ-adic Lie group*; it is not finite.

6.5 Once we have proven Thm. 4.1, one easily sees[3] that 6.1 and 6.3 *remain valid in weight 1*; however, in this case, the image of G is finite.

(b) Reduction mod ℓ

6.6 Let $K \subset \mathbb{C}$ be a number field, λ a finite prime of K, $\mathfrak{O}_\lambda$ the corresponding valuation ring, $\mathfrak{m}_\lambda$ its maximal ideal, $k_\lambda = \mathfrak{O}_\lambda/\mathfrak{m}_\lambda$ the residue field, and ℓ the characteristic of k_λ. In the sequel, we write "mod λ" for "mod $\mathfrak{m}_\lambda$."

Let f be a modular form of type (k, ε) on $\Gamma_0(N)$. One says that f is *λ-integral* (resp. $f \equiv 0 \pmod{\lambda}$) if the coefficients of the series $f_\infty(q)$ belong to $\mathfrak{O}_\lambda$ (resp. to $\mathfrak{m}_\lambda$). Suppose f is λ-integral; one says that f is an *eigenvector of $T_p \bmod \lambda$, of eigenvalue $a_p \in k_\lambda$*, if one has

$$f|T_p - a_p f \equiv 0 \pmod{\lambda}. \tag{6.6.1}$$

Theorem 6.7. *With the above notations, let f be a modular form of type (k, ε) on $\Gamma_0(N)$, $k \geq 1$, with coefficients in K. Let f be λ-integral, $f \not\equiv 0 \pmod{\lambda}$, and that f is an eigenvector of the $T_p \bmod \lambda$, for $p \nmid N\ell$, with eigenvalues $a_p \in k_\lambda$. Let k_f be the subfield of k_λ generated by the a_p and the reductions mod λ of the $\varepsilon(p)$. Then there exists a semisimple representation*

$$\rho : G \to \mathrm{GL}_2(k_f)$$

[3] This results from the fact that the representation ρ of Thm. 4.1 is *realizable over K*: its image contains a distinct rational eigenvalue (the element $\rho(c)$ of 4.5) and this leads to its Schur index being 1 (*cf.* [20], IX a); this is thus realizable in the field of values of the character.

which is unramified outside $N\ell$ and such that, for $p \nmid N\ell$, one has

$$\mathrm{Tr}(F_{\rho,p}) = a_p \quad \textit{and} \quad \det(F_{\rho,p}) \equiv \varepsilon(p)p^{k-1} \pmod{\lambda}. \tag{6.7.1}$$

Proof of Thm. 6.7

6.8 Let $(K', \lambda', f', k', \varepsilon', (a'_p))$ be as in Thm. 6.7, where K' contains K and λ' extends λ. If $a_p \equiv a'_p \pmod{\lambda'}$ and $\varepsilon(p)p^{k-1} \equiv \varepsilon'(p)p^{k'-1} \pmod{\lambda'}$ for all $p \nmid N\ell$, the theorem for f is equivalent to the theorem for f'. The second condition is verified as soon as $\varepsilon = \varepsilon'$ and $k \equiv k' \pmod{(\ell-1)}$, and with the first condition, this gives $f \equiv f' \pmod{\lambda'}$.

6.9 Reduction to the case $k \geq 2$ – For $n > 2$ even, let E_n be the Eisenstein series of weight n on $\mathrm{SL}_2(\mathbb{Z})$, normalized so that the constant term is 1. If one chooses n so that $(\ell-1) \mid n$, the power series of E_n is ℓ-entire, and $E_n \equiv 1 \pmod{\ell}$, *cf.* [25]. The product $f \cdot E_n$ is thus congruent to f mod λ; its weight $k+n$ is congruent to k mod $(\ell-1)$. By 6.8, the Thm. for f is equivalent to the Thm. for $f \cdot E_n$, which is of weight > 2.

6.10 Reduction to f an eigenfunction of the T_p – It suffices to prove the existence of an f' as in 6.8, with $(k, \varepsilon) = (k', \varepsilon')$, with f' an eigenvector of the T_p. This follows from the Lemma below, applied to the T_p acting on the $\mathfrak{O}_\lambda$-module M of modular forms of type (k, ε) on $\Gamma_0(N)$ with coefficients in $\mathfrak{O}_\lambda$:

Lemma 6.11. *Let M be a finitely generated free module on a discrete valuation ring $\mathfrak{O}$, with maximal ideal $\mathfrak{m}$, residue field k, and field of fractions K. Let $\mathcal{T}$ be a commuting family of endomorphisms of M. Let $f \in M/\mathfrak{m}M$ a nonzero common eigenvector of $T \in \mathcal{T}$, and let $a_T \in k$ be the corresponding eigenvalues. Then there exists a discrete valuation ring $\mathfrak{O}'$ containing $\mathfrak{O}$, with maximal ideal $\mathfrak{m}'$ such that $\mathfrak{O} \cap \mathfrak{m}' = \mathfrak{m}$, with field of fractions K' a finite extension of K, and a nonzero element f' of*

$$M' = \mathfrak{O}' \otimes_{\mathfrak{O}} M,$$

an eigenvector of all $T \in \mathcal{T}$ of eigenvalue a'_T, such that $a'_T \equiv a_T \pmod{\mathfrak{m}'}$. (Notice that the *eigenvectors* may not stay the same, only the *eigenvalues*.)

Let $\mathcal{H}$ be the subalgebra of $\mathrm{End}(M)$ generated by $\mathcal{T}$. After making a finite extension of scalars, one may suppose that $K \otimes \mathcal{H}$ is a product of artinian rings with residue field K. Let $\chi : \mathcal{H} \to k$ be the homomorphism such that $h \cdot f = \chi(h)f$ for all $h \in \mathcal{H}$. Since $\mathcal{H}$ is free over $\mathfrak{O}$, there exists a prime ideal $\mathfrak{p}$ of $\mathcal{H}$ contained in the maximal ideal $\ker(\chi)$ and such that $\mathfrak{p} \cap \mathfrak{O} = 0$; this is the kernel of a homomorphism $\chi' : \mathcal{H} \to \mathfrak{O}$ whose reduction mod $\mathfrak{m}$ is χ. The ideal of $K \otimes \mathcal{H}$ generated by $\mathfrak{p}$ survives in $K \otimes M$; hence we have that there exists a nonzero element f'' of $K \otimes M$ which is annihilated by this ideal, i.e. such that $hf'' = \chi'(h)f''$ for all $h \in \mathcal{H}$. One then takes for f' a nonzero multiple of f'' contained in M.

Alternate proof: One reduces to the case where M is $\mathcal{T}$-indecomposable, and where the eigenvalues of $T \in \mathcal{T}$ belong to K. One shows that there is a basis

$(e_1, \ldots, e_n)$ of M such that with respect to this basis, the elements $T \in \mathcal{T}$ are upper triangular matrices (T_{ij}); next, using the indecomposability of M, one shows that $T_{ii} \equiv a_T \pmod{\mathfrak{m}}$ for all T and for all i. The element $f' = e_1$ then has the desired properties.

6.12 End of Proof of 6.7 – By 6.9 and 6.10, we may suppose that $k \geq 2$ and that f is an eigenvector of the T_p, $p \nmid N\ell$; as T_ℓ commutes with the T_p, we may also suppose that f is an eigenvector of T_ℓ if $\ell \nmid N$. Then let

$$\rho_\lambda : G \to \mathrm{GL}_2(K_\lambda)$$

be the representation associated to f by Thm. 6.1. By replacing ρ_λ by an isomorphic representation, we may suppose that $\rho_\lambda(G)$ is contained in $\mathrm{GL}_2(\hat{\mathfrak{O}}_\lambda)$, where $\hat{\mathfrak{O}}_\lambda$ is the ring of integers of K_λ (i.e. the completion of $\mathfrak{O}_\lambda$). By reduction mod λ, one gets from ρ_λ the representation

$$\tilde{\rho}_\lambda : G \to \mathrm{GL}_2(k_\lambda).$$

Let φ be the semisimplification of $\tilde{\rho}_\lambda$; this representation is semisimple, unramified outside $N\ell$, and it satisfies (6.7.1). The group $\varphi(G)$ is finite; by the Chebotarev Density Theorem, all the elements of $\varphi(G)$ are of the form $F_{\varphi,p}$, with $p \nmid N\ell$. By the definition of k_f, one then has:

(6.12.1) For all $s \in \varphi(G)$, the coefficients of the polynomial $\det(1 - sT)$ are contained in k_f.

The existence of the representation $\rho : G \to \mathrm{GL}_2(k_f)$ we seek results from the following lemma:

Lemma 6.13. *Let $\varphi : \Phi \to \mathrm{GL}_n(k')$ be a semisimple representation of the group Φ on a finite field k'. Let k be a subfield of k' containing the coefficients of the polynomials $\det(1 - \varphi(s)T)$, $s \in \Phi$. Then φ is realizable over k, i.e. it is isomorphic to a representation $\rho : \Phi \to \mathrm{GL}_n(k)$.*

For φ to be realizable over k, it suffices to show that φ is isomorphic to $\sigma(\varphi)$ where σ is an automorphism of k' over k: this is due to the fact that the Brauer group of a finite field is trivial, and there are no "Schur indices" to consider. Now φ and $\sigma(\varphi)$ have the same characteristic polynomial, and are semisimple; hence they are isomorphic by [3], Thm. 30.16.

§7 Bounds for the orders of certain subgroups of $\mathrm{GL}_2(\mathbb{F}_\ell)$

Let ℓ be a prime number, and we denote by $\mathbb{F}_\ell$ the field of ℓ elements.

7.1 Let η and M be two positive numbers. We consider the following property of a subgroup G of $\mathrm{GL}_2(\mathbb{F}_\ell)$:

$C(\eta, M)$ – There exists a subset H of G with $|H| \geq (1 - \eta)|G|$, and such that the set of polynomials $\det(1 - hT)$, $h \in H$, has at most M elements.

(Here $|X|$ denotes the cardinality of the finite set X.)

We say that G is *semisimple* if the identity representation

$$G \to \mathrm{GL}_2(\mathbb{F}_\ell)$$

is semisimple.

Proposition 7.2. *Let $\eta < 1/2$ and $M \geq 0$. There exists a constant $A = A(\eta, M)$ such that, for every prime number ℓ, and every semisimple subgroup G of $\mathrm{GL}_2(\mathbb{F}_\ell)$ satisfying $C(\eta, M)$, we have $|G| \leq A$.*

Proof – Let G be a semisimple subgroup of $\mathrm{GL}_2(\mathbb{F}_\ell)$. Recall (*cf.* [22], §2, Prop. 15 and 16) that G satisfies one of the following conditions:

(a) G contains $\mathrm{SL}_2(\mathbb{F}_\ell)$;

(b) G is contained in a Cartan subgroup T;

(c) G is contained in the normalizer of a Cartan subgroup T, and not contained in T;

(d) the image of G in $\mathrm{PGL}_2(\mathbb{F}_\ell) = \mathrm{GL}_2(\mathbb{F}_\ell)/\mathbb{F}_\ell^\times$ is isomorphic to $\mathfrak{A}_4$, $\mathfrak{A}_5$, or $\mathfrak{S}_4$.

We have to, in each case, bound the order of G.

Case (a): Let $r = [G : \mathrm{SL}_2(\mathbb{F}_\ell)]$. One has $|G| = r\ell(\ell^2 - 1)$. On the other hand, the number of elements in $\mathrm{GL}_2(\mathbb{F}_\ell)$ with the given characteristic polynomial is $\ell^2 + \ell$, ℓ^2, or $\ell^2 - \ell$ as the polynomial in question has 2, 1, or 0 roots in $\mathbb{F}_\ell$. If G satisfies $C(\eta, M)$, one has

$$(1 - \eta)r\ell(\ell^2 - 1) = (1 - \eta)|G| \leq |H| \leq M(\ell^2 + \ell),$$

whence

$$(1 - \eta)r(\ell - 1) \leq M \quad \text{and} \quad \ell \leq 1 + \frac{M}{(1 - \eta)r} \leq 1 + \frac{M}{1 - \eta};$$

one gets a bound for ℓ in this way, and *a fortiori* a bound for $|G|$.

Case (b): At most 2 elements of T have a given characteristic polynomial. The hypothesis $C(\eta, M)$ (with $\eta < 1$) then gives

$$(1 - \eta)|G| \leq 2M,$$

giving the bound

$$|G| \leq \frac{2M}{1 - \eta}.$$

Case (c): The group $G' = G \cap T$ has index 2 in G. If G satisfies $C(\eta, M)$, G' satisfies $C(2\eta, M)$. By applying *(b)* to G', one has

$$|G| \leq \frac{4M}{1 - 2\eta}.$$

Case (d): The image of G in $\mathrm{PGL}_2(\mathbb{F}_\ell)$ is of order at most 60. The group

$$G \cap \mathrm{SL}_2(\mathbb{F}_\ell)$$

is then of order at most 120, whence G then has at most 120 elements of the given determinant, and *a fortiori* with the given characteristic polynomial. If G satisfies $C(\eta, M)$, one then has

$$(1-\eta)|G| \leq 120M,$$

which gives

$$|G| \leq \frac{120M}{1-\eta}.$$

§8 Proof of Thm. 4.1

We may suppose the modular form f is either an Eisenstein series or a cusp form.

8.1 If f is an Eisenstein series, then there are characters χ_1 and χ_2 of $\left(\mathbb{Z}/N\mathbb{Z}\right)^\times$ where $\chi_1 \cdot \chi_2 = \varepsilon$ and $a_p = \chi_1(p) + \chi_2(p)$ for $p \nmid N$ (*cf.* [9], p. 690). One then takes for ρ the reducible representation

$$\rho = \chi_1 \oplus \chi_2,$$

where the χ_i are identified with the degree one representations of G, *cf.* 4.4.

8.2 From now on, we suppose f is cuspidal. By 2.7, the a_p and $\varepsilon(p)$ are in the ring of integers $\mathfrak{O}_K$ of a number field K, which we can suppose is galois over $\mathbb{Q}$. Let L be the set of prime numbers ℓ which split completely in K. For each $\ell \in L$, choose a prime λ_ℓ of K which extends ℓ; the corresponding residue field is equal to $\mathbb{F}_\ell$. By Thm. 6.7, there exists a continuous semisimple representation

$$\rho_\ell : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{F}_\ell),$$

which is unramified outside $N\ell$, and for which

$$\det(1 - F_{\rho_\ell,p}T) \equiv 1 - a_pT + \varepsilon(p)T^2 \pmod{\lambda_\ell}$$

if $p \nmid N\ell$.

Let G_ℓ be the image of ρ_ℓ in $\mathrm{GL}_2(\mathbb{F}_\ell)$.

Lemma 8.3. *For all $\eta > 0$, there exists a constant M such that G_ℓ satisfies $C(\eta, M)$ for every $\ell \in L$.*

By Prop. 5.5, there is a subset X_η of P, the set of primes, such that $\mathrm{dens}\,.\sup X_\eta \leq \eta$ and the a_p, for $p \notin X_\eta$, form a finite set. We let $\mathcal{M}$ be

the (finite) set of polynomials $1 - a_pT + \varepsilon(p)T^2$, for $p \notin X_\eta$, and let $M = |\mathcal{M}|$. *The group G_ℓ satisfies $C(\eta, M)$ for all $\ell \in L$.* Indeed, let H_ℓ be the subset of G_ℓ formed by the Frobenius elements $F_{\rho_\ell,p}$ for $p \notin X_\eta$, and their conjugates. By the Chebotarev Density Theorem, one has that $|H_\ell| \geq (1-\eta)|G_\ell|$. On the other hand, if $h \in H_\ell$, the polynomial $\det(1 - hT)$ is the reduction $(\mathrm{mod}\, \lambda_\ell)$ of an element of $\mathcal{M}$, which belongs to a set of at most M elements. The condition $C(\eta, M)$ is hence satisfied.

Lemma 8.4. *There exists a constant A such that $|G_\ell| \leq A$ for all $\ell \in L$.*

This follows from the preceding Lemma, and Prop. 7.2.

8.5 Choose a constant A satisfying 8.4. By extending the field K (and thus decreasing L), we may suppose K contains the n^{th} roots of unity for $n \leq A$. Let Y be the set of polynomials $(1 - \alpha T)(1 - \beta T)$, where α and β are roots of unity of order $\leq A$. If $p \nmid N$, for all $\ell \in L$ with $\ell \neq p$, there exists $R(T) \in Y$ such that

$$1 - a_pT + \varepsilon(p)T^2 \equiv R(T) \pmod{\lambda_\ell}.$$

As Y is finite, there exists R such that the above congruence is satisfied for infinitely many ℓ, and hence we have the equality

$$1 - a_pT + \varepsilon(p)T^2 = R(T),$$

and we say *the polynomial $1 - a_pT + \varepsilon(p)T^2$ belongs to Y*.

8.6 Let L' be the set of $\ell \in L$ such that $\ell > A$ and for $R \neq S \in Y$, $R \not\equiv S \pmod{\lambda_\ell}$; the set $L \smallsetminus L'$ is finite, whence L' is infinite. Let $\ell \in L'$. The order of the group G_ℓ is prime to ℓ. Then, by a standard argument, the identity representation $G_\ell \to \mathrm{GL}_2(\mathbb{F}_\ell)$ is the reduction mod λ_ℓ of a representation $G_\ell \to \mathrm{GL}_2(\mathfrak{O}_{\lambda_\ell})$, where $\mathfrak{O}_{\lambda_\ell}$ is the valuation ring of λ_ℓ. By composing the latter with the canonical map $G \to G_\ell$, we obtain a representation

$$\rho : G \to \mathrm{GL}_2(\mathfrak{O}_{\lambda_\ell}).$$

By construction, ρ is unramified outside $N\ell$. If $p \nmid N\ell$, the eigenvalues of the Frobenius element $F_{\rho,p}$ are roots of unity of order $\leq A$ (since the image of ρ is isomorphic to G_ℓ, which has order $\leq A$); hence $\det(1 - F_{\rho,p}T) \in Y$. On the other hand, since the reduction of ρ mod λ_ℓ is ρ_ℓ, we have

$$\det(1 - F_{\rho,p}T) \equiv 1 - a_pT + \varepsilon(p)T^2 \pmod{\lambda_\ell}.$$

But the two polynomials $\det(1 - F_{\rho,p}T)$ and $1 - a_pT + \varepsilon(p)T^2$ are contained in Y. As these are congruent $(\mathrm{mod}\, \lambda_\ell)$, they are equal, and we have

$$\det(1 - F_{\rho,p}T) = 1 - a_pT + \varepsilon(p)T^2$$

for all $p \nmid N\ell$.

Now we replace ℓ with another prime number $\ell' \in L'$. One gets a representation $\rho' : G \to \mathrm{GL}_2(\mathfrak{O}_{\lambda_{\ell'}})$ with the same properties, but for $p \nmid N\ell'$. In particular, one has

$$\det(1 - F_{\rho,p}T) = \det(1 - F_{\rho',p}T)$$

for $p \nmid N\ell\ell'$. By Lemma 3.2, this gives that ρ and ρ' are isomorphic as representations on $\mathrm{GL}_2(K)$, and *a fortiori* as complex representations. It follows that ρ is unramified outside N, and that

$$\det(1 - F_{\rho,p}T) = 1 - a_pT + \varepsilon(p)T^2$$

for $p \nmid N$.

8.7 It remains to show that ρ is *irreducible*. If this is not the case, then it is the sum of two representations of degree 1; let their corresponding characters be χ_1 and χ_2, unramified outside N, and we have that $\chi_1\chi_2 = \varepsilon$ and

$$a_p = \chi_1(p) + \chi_2(p)$$

for $p \nmid N$. One then has

$$\sum |a_p|^2 p^{-s} = 2\sum p^{-s} + \sum \chi_1(p)\overline{\chi}_2(p)p^{-s} + \sum \chi_2(p)\overline{\chi}_1(p)p^{-s}.$$

As we let s tend to 1, we have that

$$\sum p^{-s} = \log \frac{1}{s-1} + \mathrm{O}(1).$$

On the other hand, we have $\chi_1\overline{\chi}_2 \neq 1$ (if not, one would have $\varepsilon = (\chi_1)^2$ and $\varepsilon(-1) = 1$); it then follows that (*cf.* [21], VI.4.2)

$$\sum \chi_1(p)\overline{\chi}_2(p)p^{-s} = \mathrm{O}(1) \quad \text{and} \quad \sum \chi_2(p)\overline{\chi}_1(p)p^{-s} = \mathrm{O}(1).$$

One then has

$$\sum |a_p|^2 p^{-s} = 2\log \frac{1}{s-1} + \mathrm{O}(1)$$

as $s \to 1$, which contradicts Prop. 5.1 and finishes the proof.

§9 Application to Coefficients of Modular Forms of Weight One

Let $f = \sum_{n=0}^{\infty} a_n e^{2\pi inz/M}$, $M \geq 1$, be a modular form of weight one on a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$.

(a) Bounds for $|a_n|$

Theorem 9.1. *One has $|a_n| = O(d(n))$ as $n \to \infty$.*

(Here $d(n)$ denotes the number of positive divisors of n.)

Corollary 9.2. *One has $|a_n| = O(n^\delta)$ for all $\delta > 0$.*

In fact, one knows that $d(n)$ enjoys this same property ([8], Thm. 315).

Proof of 9.1 – If $n_0 \geq 1$ is an integer, then $d(nn_0)/d(n)$ is between 1 and $d(n_0)$. It is then the same to prove the estimate (9.1) for $f(z)$ or $f(n_0 z)$, and hence we may suppose $M = 1$, i.e. $f(z+1) = f(z)$. Using 1.5 and 1.9, we reduce to one of the following two cases:

(i) f is an Eisenstein series, whence (9.1) follows from the formula for a_n ([9], p. 475);

(ii) f is a cuspidal newform of type $(1, \varepsilon)$ on $\Gamma_0(N)$, for N and ε compatible. In this case, one has the more precise result:

$$|a_n| \leq d_N(n) \leq d(n), \tag{9.3.1}$$

with $d_N(n)$ the number of positive divisors of n prime to N. In fact, as a_n and $d_N(n)$ are multiplicative, it suffices to verify (9.3.1) when n is a prime power p^m. We divide this into two cases:

(ii$_1$) $p \mid N$.

One has $a_n = (a_p)^m$, and Thm. 4.6 shows that a_p is either 0 or a root of unity. One then has

$$|a_n| \leq 1 = d_N(n).$$

(ii$_2$) $p \nmid N$.

If one writes the polynomial $1 - a_p T + \varepsilon(p)T^2$ in the form $(1-\lambda T)(1-\mu T)$, one has

$$a_n = \lambda^m + \lambda^{m-1}\mu + \cdots + \lambda\mu^{m-1} + \mu^m.$$

Also, by Thm. 4.1, λ and μ are roots of unity. One then has

$$|a_n| \leq m + 1 = d_N(n).$$

9.4 Remark – If $f = \sum b_n e^{2\pi i n z/M}$, $M \geq 1$, is a cusp form of weight $k \geq 2$ on a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, the same argument as above (using [6], 8.2) gives

$$|b_n| = \mathrm{O}(n^{(k-1)/2} d(n))$$

as $n \to \infty$.

(b) Maximal order of $|a_n|$ – One knows ([8], Thm. 317) that the "maximal" order of $d(n)$ is $2^{\log n / \log\log n}$, in the sense that

$$\limsup \frac{\log d(n) \log\log n}{\log n} = \log 2.$$

The same result holds for the $|a_n|$:

Proposition 9.5. *If $f \neq 0$, then*

$$\limsup \frac{\log |a_n| \log\log n}{\log n} = \log 2.$$

Lemma 9.6. *Let $N \geq 1$ be an integer. Then there exist sets X_N and Y_N of prime numbers, with densities > 0, such that:*

(x) For all $p \in X_N$, we have $p \equiv 1 \pmod N$ and $g|T_p = 2g$ for every modular form g of weight one on $\Gamma_1(N)$;

(y) For all $p \in Y_N$, we have $p \equiv -1 \pmod N$ and $g|T_p = 0$ for every modular form g of weight one on $\Gamma_1(N)$.

Let $\rho_1, \ldots, \rho_h$ be the representations of G associated to different systems of eigenvalues of the T_p acting on the forms of type $(1, \varepsilon)$ on $\Gamma_0(N)$, where ε runs over the odd characters of $\left(\mathbb{Z}/N\mathbb{Z}\right)^\times$. Let X_N be the set of $p \equiv 1 \pmod N$ such that $F_{\rho_i,p} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ for $i = 1, \ldots, h$, and let Y_N be the set of $p \equiv -1 \pmod N$ such that $F_{\rho_i,p}$ is conjugate to $\rho_i(c)$, *cf.* 4.5. By the Chebotarev Density Thm., X_N and Y_N have densities > 0. If $p \in X_N$, 2 is the only eigenvalue of the T_p; as T_p is semisimple, one has $g|T_p = 2g$ for all g. The same argument shows that $g|T_p = 0$ if $p \in Y_N$ as the trace of the matrix $\rho_i(c)$ is 0.

Proof of 9.5 – As in (a), one reduces to the case that f is a modular form of weight one on $\Gamma_1(N)$. Let X_N be as in Lemma 9.6, and choose an integer m such that $a_m \neq 0$. If x is an integer ≥ 1, let $p_1, \ldots, p_{i(x)}$ be the distinct prime numbers $p \in X_N$ which are $\leq x$ and do not divide m. Define $n(x) = m p_1 p_2 \cdots p_{i(x)}$. As the p_i are in X_N, one has that $f|T_{p_i} = 2f$ and $f|R_{p_i} = f$; by (2.5.1), this gives

$$a_{n(x)} = 2^{i(x)} a_m,$$

whence

$$\log |a_{n(x)}| \sim i(x) \log 2$$

as $x \to \infty$. If c is the density of X_N, one has $i(x) \sim cx/\log x$, and

$$\sum_{i \leq i(x)} \log p_i \sim cx.$$

One then has

$$\log |a_{n(x)}| \sim \frac{cx \log 2}{\log x},$$
$$\log n(x) \sim cx,$$
$$\log\log n(x) \sim \log x,$$

which gives the inequality

$$\limsup \frac{\log |a_n| \log\log n}{\log n} \geq \log 2.$$

The opposite inequality comes from the fact that $|a_n| = \mathrm{O}(d(n))$.

(c) Normal order of the $|a_n|$ – The "normal" order (i.e. the most frequent) of $d(n)$ is $2^{\log\log n}$ (*cf.* [8], Thm. 432). That for the $|a_n|$ is much smaller:

Proposition 9.7. *The set of n such that $a_n = 0$ has density 1.*

(A subset $S \subset \mathbb{N}$ is said to have density c if the number of elements of S which are $\leq x$ is equal to $cx + \mathrm{o}(x)$ as $x \to \infty$.)

Here again, we may assume that f is a modular form of weight one on $\Gamma_1(N)$. Let Y_N be as in Lemma 9.6. If $p \in Y_N$, one has that $f|T_p = 0$ and $f|R_p = -f$. As a result of 2.5.1, if n is an integer divisible by p but not p^2, one has $a_n = 0$. Yet, if Y is a finite set of prime numbers, the set S_Y of integers n with the above property (for at least one $p \in Y$) has density

$$1 - \prod_{p \in Y} \left(1 - \frac{p-1}{p^2}\right).$$

As the set Y_N has a density > 0, the series $\displaystyle\sum_{p \in Y_N} \frac{1}{p}$ diverges, and the product

$$\prod_{p \in Y_N} \left(1 - \frac{p-1}{p^2}\right)$$

hence has value zero. One then has that the union of the S_Y, $Y \subset Y_N$, is of density one, which gives the result.

9.8 Remark – For all x, let $M(x)$ be the number of $n \leq x$ such that $a_n \neq 0$. Prop. 9.7 can then be restated as

$$M(x) = \mathrm{o}(x)$$

as $x \to \infty$.

By using Thm. 2 of [23], one can prove the following more precise result: there exists $\alpha > 0$ such that

$$M(x) = \mathrm{O}\left(\frac{x}{\log^{\alpha} x}\right)$$

as $x \to \infty$.

References

[1] Artin, E. Zur theorie der L-Reihen mit allgemeinen Gruppencharakteren. *Collected Works* (1930), 165–179.

[2] Atkin, A. O. L., and Lehner, J. Hecke operators on $\Gamma_0(m)$. *Math. Ann. 185* (1970), 134–160.

[3] Curtis, C., and Reiner, I. *Representation Theory of Finite Groups and Associative Algebras*. Interscience Publishers, New York, 1962.

[4] Deligne, P. Formes modulaires et représentations ℓ-adiques. *Séminaire Bourbaki 1968/1969, exposé 355* (1971), 139–172.

[5] Deligne, P. Formes modulaires et représentations de $\mathbf{GL}(2)$. *Lecture Notes 349* (1973), 55–105.

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

[7] Deligne, P., and Rapoport, M. Les schémas de modules de courbes elliptiques. *Lecture Notes 349* (1973), 143–316.

[8] Hardy, G. H., and Wright, E. M. *Introduction to the Theory of Numbers*, 3 ed. Oxford, 1954.

[9] Hecke, E. *Mathematische Werke*. Vandenhoeck und Ruprecht, G ottingen, 1970.

[10] Jacquet, H. *Automorphic Forms on $\boldsymbol{GL}(2)$, Part II*, vol. 278 of *Lecture Notes*. Springer, 1972.

[11] Langlands, R. Modular forms and ℓ-adic representations. *Lecture Notes 349* (1973), 361–500.

[12] Li, W. Newforms and functional equations. *Math. Ann. 212* (1975), 285–315.

[13] Miyake, T. On automorphic forms on $\mathbf{GL}_2$ and Hecke operators. *Ann. of Maths. 94* (1971), 174–189.

[14] Ogg, A. *Modular Forms and Dirichlet Series in Number Theory*. W. A. Benjamin Publ., New York, 1969.

[15] Ogg, A. On a convolution of L-series. *Invent. Math. 7* (1969), 297–312.

[16] Ogg, A. On the eigenvalues of Hecke operators. *Math. Ann. 179* (1969), 101–108.

[17] Piateckii-Shapiro, I. I. Zeta Functions of Modular Curves. *Lecture Notes 349* (1973), 317–360.

[18] Rankin, R. A. Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions. I, II. *Proc. Cambridge Phil. Soc. 35* (1939), 351–372.

[19] Rankin, R. A. An Ω-result for the coefficients of cusp forms. *Math. Ann. 203* (1973), 239–250.

[20] Schur, I. Arithmetische Untersuchungen uber endliche Gruppen linearer Substitutionen. *Gesam. Abhl. I* (1973), 177–197.

[21] Serre, J.-P. *Cours d'Arithmétique*. Presses Universitaires de France, Paris, 1970.

[22] Serre, J.-P. Propriétés galoisiennes des points d'ordre fini des courbes elliptiques. *Invent. Math. 15* (1972), 259–331.

[23] Serre, J.-P. Divisibilité des coefficients des formes modularies de poids entier. *C. R. Acad. Sci. Paris t. 279, série A* (1974), 679–682.

[24] Shimura, G. *Introduction to the Arithmetic Theory of Automorphic Functions*. Princeton University Press, 1971.

[25] Swinnerton-Dyer, H. P. F. On ℓ-adic representations and congruences for coefficients of modular forms. *Lecture Notes 350* (1973), 1–55.

[26] Weil, A. Uber die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen. *Math. Ann. 168* (1967), 149–156.

[27] Weil, A. Dirichlet Series and Automorphic Functions (lezioni fermaine). *Lecture Notes in Mathematics 189* (1971), 1–55.