Lectures on Applications of Modular Forms to Number Theory

Ernst Kani
Queen's University

January 2005

Contents

Introduction

The theory of (elliptic) modular forms was developed in the 19th century by Klein, Fricke, Poincaré, Weber and others, building upon the theory of elliptic functions which had evolved from the earlier work of Euler, Jacobi, Eisenstein, Riemann, Weierstrass and many others. Thus, from the outset, modular functions were intimately linked to the study of elliptic functions/elliptic curves and this relation between these theories has remained throughout its development for the benefit of both.

The basic fascination of modular forms may be summarized as follows.

1. The basic concepts of modular forms are extremely simple and require vitually no technical preparation (as we shall see). Nevertheless, it is a subject in which many diverse areas of mathematics are fused together:

   - complex analysis, in particular Riemann surfaces
   - algebra, algebraic geometry
   - (non-euclidean) geometry
   - (matrix) group theory, Lie groups, representation theory
   - number theory, arithmetic algebraic geometry

   This interaction goes in both directions: on the one hand, the above areas supply the tools necessary for solving many of the problems studied in the theory of modular forms; on the other hand, the latter furnishes important explicit examples which not only illustrate and illuminate but only advance the general theory of many of these branches.

2. Many of the basic results in the theory are very explicit and hence suitable for computational purposes, be it by hand or by computer.

3. One of the most fascinating aspects of modular forms and functions is the universality of their applications, not only in number theory but also in many other branches of Mathematics. This is in part due to the fact that modular forms are functions "with many hidden symmetries", and such functions naturally arise in many applications, even in Physics. Some of these include:

- Analysis: *Ruziewiecz's problem* — the uniqueness of finitely additive measures on $S^n, n \geq 2$.
  (Margulis (1980), Sullivan (1981), Drinfeld (1984), Sarnak (1990); cf. Sarnak [Sa])
- Algebraic topology: elliptic genera — spin manifolds, representations of the cobordism ring.
  (Witten (1983), Landweber, Stong, Ochanine, Kreck (1984ff); cf. Landweber [Land])
- Lie algebras (group theory): Kac-Moody algebras — connections with the Dedekind η-function
  (Kac, Moody, MacDonald [Mc] (1972ff)); conjectural relations with the Monster group ("Monstrous Moonshine" — Conway/Norton[CN], 1979).
- Graph theory, telephone network theory: the construction of expander graphs.
  (Alon (1986), Lubotzky/Phillips/Sarnak (1986ff); cf. Bien[Bi], Sarnak[Sa])
- Physics: string theory (cf. elliptic genera, moduli theory etc.)

The applications of modular forms to number theory are legion; in fact, as Sarnak says in his book[Sa], "traditionally the theory of modular forms has been and still is, one of the most powerful tools in number theory". Some of these applications are the following:

- Elementary number theory: identities for certain arithmetic functions.
  (Jacobi (1830), Glaisher (1885), Ramanujan (1916), ...)
- Analytic number theory:
  - Orders of magnitude of certain functions.
    (Ramanujan (1916), Hardy (1920), ...)
  - Dirichlet series, Euler products and functional equations.
    (Hecke(1936), Weil(1967), Atkin-Lehner(1970), Li(1972) ...)
- Algebraic number theory:
  - Complex multiplication *"Kronecker's Jugendtraum";* cf. Hilbert's 12^{th} problem — the generation of class fields of $\mathbb{Q}(\sqrt{-D})$.
    (Weber (1908), Fueter (1924), Hasse (1927), Deuring (1947); cf. Borel[Bo])
  - The arithmetic of positive definite quadratic forms — formulae, relations, the order of magnitude of the number of representations.
    (Hecke, Siegel, 1930ff)
  - The Gauss conjecture on class numbers of imaginary quadratic fields.
    (Heegner (1954), Goldfeld, Gross/Zagier (1983ff))
  - Two-dimensional Galois representations of $\mathbb{Q}$ and Artin's conjecture on L-functions.
    (Weil, Langlands (1971), Deligne/Serre (1974))

- The arithmetic of elliptic curves.
  (Tate, Mazur, Birch, Swinnerton-Dyer, Serre, Wiles, ... (1970ff))
- The congruent numbers problem.
  (Tunnel, 1983; cf. Koblitz[Ko])
- Fermat's Last Theorem
  (Frey, Serre, Mazur, Ribet (1987ff), Wiles (1995))

4. It lies at the fore-front of present-day mathematical research. This is not only due to its many deep applications as mentioned above, but also because it is a stepping stone for a number of other mathematical research areas which have experienced tremendous growth in the last few decades, such as:

   - The theory of automorphic forms (Shimura, Langlands)
   - Langland's program: representation theory of adele groups
     (Jacquet, Langlands, Kottwitz, Closzel, Arthur)
   - Hibert modular forms (Hirzebruch, Zagier, van der Geer)
   - Siegel modular forms and moduli of abelian varieties
     (Mumford, Deligne, Faltings, Chai).

Many of the above applications of modular forms are based on the following simple idea. Suppose we are given a sequence $A = \{a_n\}_{n\geq 1}$ of real or complex numbers whose behaviour we want to understand. Consider the associated "generating function"

$$f_A(z) = a_0 + \sum_{n\geq 1} a_n q^n, \quad \text{in which } q = e^{2\pi i z}$$

(and a_0 is chosen "suitably"). If the a_n's do not grow too rapidly, then this sum converges for all z in the upper half plane $\mathfrak{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$, and hence $f_A(z)$ is a holomorphic (= complex-differentiable) function on $\mathfrak{H}$. Clearly, f_A is invariant under translation (by 1), i.e.

$$f_A(z+1) = f_A(z),$$

and hence f_A has a *built-in* symmetry. If it also has *other* (hidden) symmetries, for example, if

$$f_A(-1/z) = z^k f_A(z), \quad \forall z \in \mathfrak{H},$$

for some k, then f_A is called a *modular form of weight k* (provided that a certain technical condition holds).

Now if f_A is a modular form of weight k, then it is determined by its first $m+1$ (Fourier) coefficients $a_0, \ldots, a_m$ where $m = [\frac{k}{12}]$, i.e. by a *finite set of data*; cf. Corollary 1.4. In particular, the space of modular forms of fixed weight k is a finite-dimensional $\mathbb{C}$-vector space, and any linear relation among the first Fourier coefficients of modular forms *holds universally.* This is the basis of many of the applications, particularly those which establish *identities* between f_A and other (known) modular forms.

For example, consider the case that

$$a_n = \sigma_{k-1}(n) := \sum_{d|n} d^{k-1}$$

is the sum of $(k-1)$st powers of the (positive) divisors of n. If k is even and $k > 2$, then (for suitable constant $a_0 = a_{0,k}$) the function

$$E_k = a_0 + \sum_{n \geq 1} \sigma_{k-1}(n) q^n$$

is a modular form of weight k. In particular, we see that E_4^2 and E_8 are both modular forms of weight 8, so by comparing the constant coefficients it follows that $E_4^2 = E_8$. We thus obtain the curious identity

$$120 \sum_{k=1}^{n} \sigma_3(k)\sigma_3(n-k) = \sigma_7(n) - \sigma_3(n), \quad n \geq 1,$$

and other identities are derived in a similar manner; cf. subsection 1.3.3.

The purpose of these lectures is to give a rough outline of some of the aforementioned applications of modular forms to number theory. Since no prior knowledge of modular forms is presupposed, the basic definitions and results of the theory are surveyed in some detail, but mainly without proofs. The latter may be found in standard texts such as Serre[Se1], Koblitz[Ko], Schoeneberg[Sch], Iwaniec[Iw], etc.

Chapter 1

Modular Forms on $\mathbf{SL}_2(\mathbb{Z})$

1.1 The definition of modular forms and functions

Modular functions are certain functions defined on the upper half-plane

$$\mathfrak{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$$

which are invariant, or almost invariant, with respect to a subgroup $\Gamma \subset \Gamma(1)$ of the modular group $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$. Here $\Gamma(1)$ or, more generally, the group

$$G = \mathrm{GL}_2^+(\mathbb{R}) = \{g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) : a, b, c, d \in \mathbb{R},\ \det(g) > 0\}$$

operates on $\mathfrak{H}$ via fractional linear transformations:

$$g(z) = \frac{az+b}{cz+d}, \quad \text{if} \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{1.1}$$

To make this more precise, let us first introduce the following preliminary concept.

Definition. Let $k \in \mathbb{Z}$. We say that a function f is *weakly modular of weight k on* Γ (or: *with respect to* Γ) if

1) f is meromorphic on $\mathfrak{H}$;

2) f satisfies the transformation law

$$f(g(z)) = j(g,z)^k f(z), \quad \forall g \in \Gamma, \tag{1.2}$$

in which $j(g,z) = cz + d$ if $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$.

Remarks. 0) Recall from complex analysis (cf. e.g. [Ah], p. 128) that a function f defined on an open set $U \subset \mathbb{C}$ is called *meromorphic* if it has for every $a \in U$ a *Laurent expansion*

$$f(z) = \sum_{n=n_a}^{\infty} c_{n,a}(z-a)^n$$

which converges in a (punctured) neighbourhood of a. (If the integer n_a can be chosen to be non-negative, then f is said to be *holomorphic* or *analytic* at a.)

1) The above functional equation (1.2) may be written in a more convenient form if we introduce the operator $|_k g$ (or $|[g]_k$):

$$f(z)|_k g := f(z)|[g]_k := f(g(z))j(g,z)^{-k}, \quad \text{for } g \in \mathrm{SL}_2(\mathbb{Z}).$$

Indeed, by using this operator, we can then write equation (1.2) in the equivalent form

$$f|_k g = f, \quad \forall g \in \Gamma. \tag{1.3}$$

It is useful to observe that the operator $|_k g$ satisfies the "associative law"

$$f|_k(g_1 g_2) = (f|_k g_1)|_k g_2, \quad \text{for all } g_1, g_2 \in \mathrm{SL}_2(\mathbb{Z}); \tag{1.4}$$

this follows immediately from the following "cocycle condition" (which is easily verified):

$$j(g_1 g_2, z) = j(g_1, g_2(z))j(g_2, z).$$

Note that it follows from (1.4) that for a fixed f (and k), the set of $g \in \mathrm{GL}_2^+(\mathbb{R})$ which satisfy (1.2) (or, equivalently, (1.3)) is a subgroup of $\mathrm{GL}_2^+(\mathbb{R})$. Thus every meromorphic f on $\mathfrak{H}$ is weakly modular of weight k for *some* subgroup $\Gamma \leq \mathrm{GL}_2^+(\mathbb{R})$.

2) Note that since we have $f|_k(-1) = (-1)^k f$, it follows that if k is odd and $-1 \in \Gamma$, then there is no weakly modular function of weight k on Γ other than the function 0. In particular, there are no non-zero weakly modular forms of odd weight on $\Gamma(1)$.

3) For later reference let us also observe that

$$j(g,z) = 1, \forall z \;\Leftrightarrow\; g \in \Gamma_\infty := \{\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) : n \in \mathbb{Z}\} \;\leq\; \Gamma(1).$$

Thus, by using the transformation law (1.4) we see that

$$j(g_1,z) = j(g_2,z), \forall z \quad \Leftrightarrow \quad g_1 \in \Gamma_\infty g_2.$$

We now come to the definition of a modular function on Γ: this is a weakly modular function on Γ which satisfies an extra condition. Since the formulation of this condition for an arbitrary subgroup is somewhat more involved, we shall focus for the moment on the case that $\Gamma = \Gamma(1)$ and treat the more general case in a later chapter; cf. Ch. 2.

Since $\Gamma = \Gamma(1)$ contains the matrices $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, we see that condition (1.2) above implies

$$f(z+1) = f(z), \tag{1.5}$$

$$f(-1/z) = z^k f(z). \tag{1.6}$$

(In fact, since T and S generate $\Gamma(1)$ (cf. Serre[Se1], p. 78), it follows that properties (1.5) and (1.6) are actually equivalent to property (1.2).)

Now condition (1.5) means that f is a periodic function (of period 1), and so f has a Fourier expansion

$$f(x) = \sum_{n=-\infty}^{\infty} a_n q^n, \quad \text{where } q = \exp(2\pi i z).$$

We say that f is *meromorphic at* ∞ if we have $a_n = 0$ for $n \leq -n_0$ for some n_0, and that f is *holomorphic at* ∞ if $a_n = 0$ for $n < 0$. Moreover, f is said to *vanish at* ∞ if $a_n = 0$ for $n \leq 0$.

Definition. (a) A *modular function of weight* k on $\Gamma = \Gamma(1)$ is a weakly modular function of weight k on Γ which is meromorphic at ∞.

(b) A *modular form* of weight k on Γ is a modular function which is holomorphic on $\mathfrak{H}$ and at ∞.

(c) A *cusp form* (*Spitzenform* in German) is a modular form which vanishes at ∞.

Notation: Let

$$\begin{array}{lcll} \mathbf{A}_k & = & \mathbf{A}_k(\Gamma) & \text{denote the space of modular functions of weight } k \text{ on } \Gamma, \\ \mathbf{M}_k & = & \mathbf{M}_k(\Gamma) & \text{denote the space of modular forms of weight } k \text{ on } \Gamma, \\ \mathbf{S}_k & = & \mathbf{S}_k(\Gamma) & \text{denote the space of cusp forms of weight } k \text{ on } \Gamma. \end{array}$$

We thus have the inclusions $\mathbf{S}_k \subset \mathbf{M}_k \subset \mathbf{A}_k$.

Remark. It is clear that $\mathbf{A}_k$, $\mathbf{M}_k$, and $\mathbf{S}_k$ are $\mathbb{C}$-vector spaces, and it is not difficult to see that

$$\begin{array}{rcccll} \mathbf{A} & = & \sum \mathbf{A}_k & = & \oplus \mathbf{A}_k & \text{is a graded field,} \\ \mathbf{M} & = & \sum \mathbf{M}_k & = & \oplus \mathbf{M}_k & \text{is a graded ring, and} \\ \mathbf{S} & = & \sum \mathbf{S}_k & = & \oplus \mathbf{S}_k & \text{is a graded ideal of } \mathbf{M}, \end{array}$$

where the above sums are over all $k \in \mathbb{Z}$ and are taken in $\mathcal{M}(\mathfrak{H})$, the field of all meromorphic functions on $\mathfrak{H}$. Note that the functions in $\mathbf{A}$ (etc.) no longer satisfy a transformation law with respect to $\Gamma(1)$; this is analogous to the fact that the sum of two or more eigenvectors associated to different eigenvalues are in general longer eigenvectors.

Nevertheless, it is useful to study these large (abstract) spaces since it turns out that they have a relatively simple structure: $\mathbf{M}$ is a graded polynomial ring in two variables, and $\mathbf{S}$ is a principal $\mathbf{M}$-ideal; cf. Theorem 1.1. In particular, it follows that $\mathbf{M}_k$ and $\mathbf{S}_k$ are finite-dimensional vector spaces (for all $k \in \mathbb{Z}$).

1.2 Examples

Before continuing with the general theory of modular forms, let us look at some basic examples.

1.2.1 Eisenstein series

These are the series defined by

$$G_k(z) \;=\; {\sum_{m,n\in\mathbb{Z}}}' \frac{1}{(mz+n)^k}$$

which converge absolutely for $k \geq 3$. Here the prime on the summation sign indicates that term $(m,n) = (0,0)$ has been omitted. Note that we can also write this sum as

$$G_k(z) = \sum_{d=1}^{\infty} \sum_{\substack{m,\,n \\ (m,\,n)\,=\,d}} \frac{1}{(mz+n)^k} = \sum_{d=1}^{\infty} \frac{1}{d^k} \sum_{\substack{m,\,n \\ (m,\,n)\,=\,1}} \frac{1}{(mz+n)^k} = \zeta(k) \sum_{\substack{m,\,n \\ (m,\,n)\,=\,1}} \frac{1}{(mz+n)^k},$$

where $\zeta(k) = \sum n^{-k}$ denotes the Riemann ζ-function. Now since every pair (m,n) with $\gcd(m,n) = 1$ can be completed to matrix $g = \left(\begin{smallmatrix} a & b \\ m & n \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, and since g is unique up left multiplication by $T^n \in \Gamma_\infty$ (cf. Remark 3) above), we can also write this as

$$G_k(z) = \zeta(k) \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} \frac{1}{j(\gamma,z)^k} = \zeta(k) \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} 1_{|_k}\gamma. \tag{1.7}$$

Facts. 0) $G_k = 0$ for $k \equiv 1(2)$.

1) G_k is a modular form of weight k for $k \geq 3$, i.e. $G_k \in \mathbf{M}_k$.

2) For $k \equiv 0\,(2)$, the q-expansion of G_k is:

$$G_k(z) = 2\zeta(k)E_k(z),$$

where $\zeta(s) = \sum n^{-s}$ is the Riemann zeta-function and

$$E_k(z) = 1 + c_k \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n. \tag{1.8}$$

Here $\sigma_{k-1}(n)$ is the sum of the k-1st powers of all divisors of n, i.e.

$$\sigma_{k-1}(n) \;=\; \sum_{d|n} d^{k-1}, \quad \text{and} \quad c_k \;=\; -\frac{2k}{B_k} \overset{\text{Euler}}{=} \frac{(2\pi i)^k}{(k-1)!\zeta(k)},$$

where B_k denotes the k^{th} Bernoulli number defined by

$$\sum_{k=0}^{\infty} B_k \frac{z^k}{k!} = \frac{z}{e^z - 1}.$$

For later reference, let us make a table of the values of c_k for small values of k:

k	2	4	6	8	10	12	14	16	18	20
c_k	-24	240	-504	480	-264	$\frac{65520}{691}$	-24	$\frac{16320}{3617}$	$-\frac{28728}{43867}$	$\frac{13200}{174611}$

Thus, the q-expansions of first two (non-zero) Eisenstein series are

$$E_4(z) = 1 + 240\sum_{n=1}^{\infty}\sigma_3(n)q^n \quad \text{and} \quad E_6(z) = 1 - 504\sum_{n=1}^{\infty}\sigma_5(n)q^n. \tag{1.9}$$

Remarks. 1) Whereas Facts 0) and 1) are clear from the definitions (and formula (1.4)), Fact 2) requires a bit more work; cf. Serre[Se1], p. 92, Schoeneberg[Sch], p. 55 or Koblitz[Ko], p. 110. (Note that Serre writes $E_{k/2}$ and $G_{k/2}$ in place of E_k and G_k.) For example, Serre and Koblitz derive (1.8) by using certain expansions of the cotan(πz) function to obtain the relation

$$\sum_{m\in\mathbb{Z}}\frac{1}{(m+z)^k} = \frac{(-2\pi i)^k}{(k-1)!}\sum_{n=1}^{\infty}n^{k-1}q^n,$$

from which (1.8) follows readily.

2) The Eisenstein series are the special case $m = 0$ of the *Poincaré series* $P_{m,k}$ defined by

$$P_{m,k}(z) = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma}\frac{1}{j(\gamma,z)^k}\exp(2\pi i m\gamma(z)).$$

For $m > 0$ and $k \geq 3$ the Poincaré series are cusp forms of weight k whose Fourier expansion may be expressed in terms of Kloosterman sums and Bessel functions. (For further information, cf. Gunning[Gu], Miyake[Mi].)

3) For $k = 2$ the infinite sum in the definition of G_k still converges but not absolutely. However, if we define

$$G_2(z) = \sum_{m=-\infty}^{\infty}{\sum_{n=-\infty}^{\infty}}'\frac{1}{(mz+n)^2} = 2\zeta(2) + \sum_{m=1}^{\infty}\sum_{n=-\infty}^{\infty}\frac{1}{(mz+n)^2} = 2\zeta(2)E_2(z),$$

then G_2 and E_2 are holomorphic functions on $\mathfrak{H}$ and we have, similar to before,

$$E_2(z) = 1 - 24\sum_{m=1}^{\infty}\sigma(n)q^n, \tag{1.10}$$

so E_2 is also holomorphic at ∞. However, E_2 isn't a modular form since it satisfies the transformation law

$$E_2(-1/z) = z^2E_2(z) + \frac{12z}{2\pi i}; \tag{1.11}$$

cf. [Ko], p. 113. (E_2 is sometimes called a *quasi-modular form.*) Nevertheless, E_2 is useful in constructing modular forms because we have the following observation which Ramanujan[Ra] made in 1916:

$$f \in \mathbf{M}_k \Longrightarrow \theta f - \frac{k}{12}fE_2 \in \mathbf{M}_{k+2}, \tag{1.12}$$

where θ denotes the derivative operator

$$\theta f = \frac{1}{2\pi i}\frac{df}{dz} = q\frac{df}{dq} = \sum na_nq^n, \quad \text{if } f = \sum a_nq^n. \tag{1.13}$$

1.2.2 The discriminant form

Following time-honoured tradition, put

$$\begin{aligned} g_2 &= 60G_4 &&= \tfrac{4\pi^4}{3}E_4, \\ g_3 &= 140G_6 &&= \tfrac{8\pi^6}{27}E_6, \\ \Delta &= g_2^3 - 27g_3^2 &&= \tfrac{(2\pi)^{12}}{1728}(E_4^3 - E_6^2). \end{aligned}$$

It is immediate from its definition that Δ is a modular form of weight 12, and a more careful analysis shows that $\Delta \neq 0$ in $\mathfrak{H}$; cf. (1.39) below. Furthermore, the q-expansions for the E_k's show that Δ vanishes at ∞, so Δ is a cusp form of weight 12, i.e. $\Delta \in \mathbf{S}_{12}$. Let us write

$$\Delta(z) = (2\pi)^{12} \sum_{n \geq 1} \tau(n) q^n;$$

this defines the *Ramanujan function* $\tau(n)$, which has been studied extensively in the literature.

Remarks. 1) We have $\tau(n) \in \mathbb{Z}, \forall n \in \mathbb{Z}$; cf. subsection 1.2.4 below. (It is clear from the definition that $\tau(n) \in \frac{1}{1728}\mathbb{Z} \subset \mathbb{Q}$ because E_4 and E_6 have integral q-expansions.)

2) In 1916 Ramanujan[Ra] showed that

$$\tau(n) \equiv \sigma_{11}(n) \pmod{691}$$

(cf. Corollary 1.9 below), and other congruences were found in subsequent years. By studying the associated ℓ-adic representation (à la Deligne, Serre) and the theory of modular forms mod p, Swinnerton-Dyer was able to show that all possible congruences have now been found; cf. [SwD].

3) Ramanujan also made a number of conjectures about $\tau(n)$:

$$\tau(nm) = \tau(n)\tau(m), \text{ if } (n,m) = 1; \tag{1.14}$$

$$\tau(p^{n+1}) = \tau(p)\tau(p^n) - p^n\tau(p^{n-1}), \text{ if } n > 1 \text{ and } p \text{ is prime}; \tag{1.15}$$

$$|\tau(p)| \leq 2p^{11/2}, \text{ if } p \text{ is prime}. \tag{1.16}$$

Of these, (1.14) and (1.15) were first proven by Mordell (1917), but now follow more easily from the formalism of Hecke operators developed by Hecke in the 1930's, as we shall see in section 1.5The third conjecture is much deeper. Deligne[De1] showed in 1968 that one can deduce this (non-trivially!) from the general *Weil Conjectures*, which he then subsequently proved in 1974 (cf. Deligne[De2]).

4) The following question proposed by D.H. Lehmer is still open:

$$\tau(n) \neq 0, \quad \text{for all } \ n \geq 1?$$

Some partial results in this direction (which are also valid for more general modular forms) were obtained by Serre; cf. Serre[Se2], §7.6.

1.2.3 The j − invariant

This is the modular function (of weight 0) defined by

$$j(z) = 1728\frac{g_2^3}{\Delta} = 1728\frac{E_4^3}{E_4^3 - E_6^2}. \tag{1.17}$$

It is holomorphic in $\mathfrak{H}$ (because $\Delta \neq 0$) and has a simple pole at ∞ with q-expansion

$$j(z) = \frac{1}{q} + 744 + \sum_{n\geq 1} c(n)q^n.$$

Remarks. 1) One can show that the Fourier coefficients $c(n)$ are integral; for example,

$$c(1) = 196884 = 2^2 3^3 1823, \qquad c(2) = 21493760 = 2^{11} \cdot 5 \cdot 2099.$$

2) One has the following congruences for the Fourier coefficients $c(n)$:

$$n \equiv 0 \pmod{p^a} \quad \Rightarrow \quad c(n) \equiv 0 \pmod{p^a}, \text{ for } a \geq 1,\ p \leq 11,\ p \text{ prime},$$

and even stronger congruences are valid for $p \leq 5$; cf. Serre[Se1], p. 90.

3) Based on observations of J. Thompson and J. McKay, Conway and Norton [CN] have advanced the conjecture that the coefficients $c(n)$ are simple linear combinations of the degrees of the irreducible characters of the "Monster group" M which is a simple group of order $2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71$.

1.2.4 The Dedekind η-function

Consider the function $\eta(z)$ defined by

$$\eta(z) = e^{2\pi iz/24}\prod_{n=1}^{\infty}(1 - e^{2\pi iz}). \tag{1.18}$$

This function is closely related to both E_2 and to the discriminant function Δ. First of all, its logarithmic derivative is

$$\frac{\eta'(z)}{\eta(z)} = \frac{2\pi i}{24}E_2(z), \tag{1.19}$$

as is easy to see (cf. [Ko], p. 121). From this and (1.11) one deduces that $\eta(z)$ satisfies the transformation laws

$$\eta(z+1) = e^{2\pi i/24}\eta(z), \tag{1.20}$$

$$\eta(-1/z) = \left(\frac{z}{i}\right)^{1/2}\eta(z), \tag{1.21}$$

which show in particular that η is not a modular form on $\Gamma(1)$. (It can, however, be considered as a modular form of weight $\frac{1}{2}$ on a subgroup of $\Gamma(1)$, as we shall see later.)

In addition, the above formulae show that its 24th power η^{24} does satisfy the transformation rules (1.5) and (1.6) with $k = 12$, and so η^{24} is a cusp form of weight 12 on $\Gamma(1)$; in fact, we have

$$\Delta(z) = (2\pi)^{12}\eta^{24}(z) = (2\pi)^{12} q \prod_{n=1}^{\infty}(1-q^n)^{24}, \tag{1.22}$$

which is a formula that was first established by Jacobi. Note that it follows from this formula that the n-th Fourier coefficient of η^{24} is given by the Ramanujan function $\tau(n)$; in particular, the $\tau(n)$'s are integral, as promised.

Remarks. 1) The η-function is also related to the *partition function* $p(n)$ via the relation

$$\frac{e^{2\pi i z/24}}{\eta(z)} = \sum_{n=0}^{\infty} p(n) q^n. \tag{1.23}$$

(Recall that the $p(n)$ denotes the number of partitions $n = n_1 + \cdots + n_s$ of n into positive integers n_i with order disregarded.) For more information about the η-function and its application to the partition function $p(n)$, cf. Knopp[Kn].

2) From (1.20) and (1.21) it follows easily that the η-function satisfies the general transformation law

$$\eta(g(z)) = c(g) j(g,z)^{\frac{1}{2}} \eta(z), \quad \text{for } g \in \mathrm{SL}_2(\mathbb{Z}), \tag{1.24}$$

for some constant $c(g) \in \mathbb{C}$ (because S and T generate the group $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$, as was mentioned earlier). However, the explicit determination of the constant $c(g)$ in terms of g is rather complicated and was first done by Dedekind: if $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $c > 0$, then

$$c(g) = \exp\left(\frac{a+d-3c}{24c} - \frac{1}{2}s(d,c)\right), \text{ where } s(d,c) = \sum_{0 \le n < c} \frac{n}{c}\left(\left(\frac{dn}{c}\right)\right)$$

denotes the so-called *Dedekind sum* in which $((x)) = x - [x] - \frac{1}{2}$; cf. Iwaniec[Iw], p. 45 and Lang[La], ch. IX for more details.

3) The Fourier expansion of η (in terms of $q^{\frac{1}{24}}$) is given by the formula

$$\eta(z) = \sum_{n=-\infty}^{\infty} (-1)^n q^{(1+12n(3n+1))/24} = \sum_{\substack{n \ge 1 \\ n \equiv \pm 1 (12)}} q^{n^2/24} - \sum_{\substack{n \ge 1 \\ n \equiv \pm 5 (12)}} q^{n^2/24},$$

which is (essentially) a famous identity due to Euler; cf. Hardy-Wright[HW], p. 284 and Iwaniec[Iw], p. 45.

1.2.5 Theta series

Another common source of modular forms is via theta series attached to quadratic forms; these are of fundamental interest in many applications. To define these, let

$$Q(x_1, \ldots, x_r) \;=\; \frac{1}{2}\sum_{i,j=1}^{r} a_{ij}x_i x_j \;=\; \sum_{1\le i\le j\le r} b_{ij}x_i x_j$$

be an even, integral, positive definite quadratic form in r variables. This means:

- The matrix $A = (a_{ij})$ is symmetric, with integral entries $a_{ij} \in \mathbb{Z}$ and even diagonal entries ($a_{ii} \in 2\mathbb{Z}$), or equivalently, the matrix $B = (b_{ij})$ defined by $b_{ii} = \frac{1}{2}a_{ii}, b_{ij} = a_{ij},$ if $i \neq j$ is integral;
- we have $Q(\vec{x}) := Q(x_1, x_2, \ldots, x_n) = \frac{1}{2}\vec{x}^t A\vec{x} > 0$, for all $\vec{x} = (x_1, \ldots, x_r) \neq \vec{0}$.

Example. If $r = 2$, then each even integral (binary) quadratic form can be written as

$$Q(x_1, x_2) = ax_1^2 + bx_1x_2 + cx_2^2 = \frac{1}{2}\vec{x}^t A\vec{x} \quad \text{where } A = \begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix} \text{ with } a, b, c \in \mathbb{Z}.$$

By completing the square, we obtain $Q(x_1, x_2) = a\left(x_1 + \frac{b}{2a}x_2\right)^2 + \frac{\det(A)}{2a}x_2^2$, where $\det(A) = 4ac - b^2$, and so we see that

$$Q = Q_A \text{ is } \textit{positive definite} \text{ if and only if } a > 0 \text{ and } \det(A) = 4ac - b^2 > 0.$$

A fundamental question, which was responsible for the development of much of Number Theory (Diophantus, Fermat, Euler, Lagrange, Gauss, ...), is the following:

Problem. Given an even integral positive definite quadratic form Q and an integer $n \geq 1$, determine the *number of representations of n by Q*, i.e. the number

$$r_Q(n) = \#\{\vec{m} \in \mathbb{Z}^r : Q(\vec{m}) = n\}.$$

As was mentioned in the introduction, one method to study a sequence of numbers is to consider its associated generating function. For this, consider the *theta series* associated to Q or to A which is defined by

$$\vartheta_Q(z) = \sum_{\vec{m}\in\mathbb{Z}^r} q^{Q(\vec{m})} = \sum_{\vec{m}\in\mathbb{Z}^r} e^{\pi i z \vec{m}^t A\vec{m}}, \tag{1.25}$$

where as usual $q = e^{2\pi iz}$. It is immediate that we have

$$\vartheta_Q(z) \;=\; \sum_{n=0}^{\infty} r_Q(n)q^n \;=\; 1 + \sum_{n=1}^{\infty} r_Q(n)q^n, \tag{1.26}$$

so ϑ_Q is indeed the generating function of the $r_Q(n)$'s. Since this sum converges on all of $\mathfrak{H}$ (cf. [Sch], p. 204 or [Se1], p. 108), it follows that ϑ_Q is a holomorphic function on $\mathfrak{H}$.

From equation (1.26) it is clear that ϑ_Q satisfies the transformation law

$$\vartheta_Q(z+1) = \vartheta_Q(z). \tag{1.27}$$

However, ϑ_Q will not be a modular form for the full modular group $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ unless Q is *unimodular*, i.e. unless its determinant

$$\det(Q) \stackrel{\text{def}}{=} \det(A) = 1.$$

In this case we have the transformation law

$$\vartheta_Q\left(-\frac{1}{z}\right) = (iz)^{r/2}\vartheta_Q(z), \tag{1.28}$$

which one can prove using the *Poisson summation formula*; cf. [Se1], p. 109. From this one easily concludes the following useful fact:

> If $Q(x_1, \ldots, x_r)$ is an even, integral, positive definite unimodular quadratic form in r variables, then $r \equiv 0 \pmod 8$.

Indeed, if false, then by replacing Q by $Q \oplus Q$ or by $Q \oplus Q \oplus Q \oplus Q$, we may assume that $r \equiv 4 \pmod 8$ Then by equations (1.27) and (1.28) we have, using the notation of section 1.1,

$$\vartheta_Q|[ST]_{\frac{r}{2}} \stackrel{(1.4)}{=} (\vartheta_Q|[S]_{\frac{r}{2}})|[T]_{\frac{r}{2}} \stackrel{(1.28)}{=} -\vartheta_Q|[T]_{\frac{r}{2}} \stackrel{(1.27)}{=} -\vartheta_Q,$$

which yields a contradiction since $(ST)^3 = 1$. Thus, $r \equiv 0 \pmod 8$.

Therefore, equation (1.28) reduces to

$$\vartheta_Q\left(-\frac{1}{z}\right) = z^{r/2}\vartheta_Q(z), \tag{1.29}$$

which, together with (1.27) and the q-expansion (1.26), implies that ϑ_Q is a modular form of weight $r/2$, i.e.

$$\vartheta_Q(z) \in \mathbf{M}_{r/2}(\Gamma), \quad \text{if } Q \text{ is unimodular and } r \equiv 0 \pmod 8. \tag{1.30}$$

Example. Consider the 8×8 matrix

$$A = \begin{pmatrix} 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix},$$

whose associated quadratic form $Q_A(x_1,\dots,x_8) = \frac{1}{2}\vec{x}^t A\vec{x}$ is given by

$$\begin{aligned} Q_A(x_1,\dots,x_8) &= x_1^2+x_2^2+x_3^2+x_4^2+x_5^2+x_6^2+x_7^2+x_8^2 \\ &\quad -x_1x_3-x_2x_4-x_3x_4-x_4x_5-x_5x_6-x_6x_7-x_7x_8. \end{aligned}$$

It is immediate that A is even and symmetric. To see that $A=(a_{ij})$ is positive definite, it is enough to verify that the principal subdeterminants $\det(A_k) = \det((a_{ij})_{1\le i,j\le k})$ are positive for $1 \le k \le 8$:

$$d_1 = \det(A_1) = 2, \quad d_2 = \det\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = 4, \quad d_3 = \det\begin{pmatrix} 2 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix} = 6,$$

and similarly, $d_4=5$, $d_5=4$, $d_6=3$, $d_7=2$, $d_8=1$. Thus A is positive definite and unimodular, and hence by (1.30), the associated theta series is a modular form of weight $4=\frac{8}{2}$ on $\Gamma(1)$, i.e. $\vartheta_A \in \mathbf{M}_4$. In fact, we shall prove later that

$$\vartheta_A = E_4,$$

which means equivalently that the number of representations of a number n by Q_A is given by the formula

$$r_{Q_A}(n) = 240\sigma_3(n), \quad \text{for } n \ge 1.$$

(To see that this is an equivalent formulation of the previous equation, use equations (1.26) and (1.9).)

Remarks. 1) As is explained in Serre[Se1], p. 51, the lattice associated to the quadratic form of the above example is the root lattice of type E_8 which arises in the theory of Lie groups. Thus, we see that the E_8-lattice has precisely $240\sigma_3(n)$ vectors of length $\sqrt{2n}$.

2) If r is even but A is not unimodular, then ϑ_A turns out to be a modular form of weight $r/2$ for some suitable subgroup $\Gamma \le \Gamma(1)$, as we shall see later. For example, if $Q = Q_A(x_1,x_2)$ is a positive definite binary quadratic form of determinant $N = \det(A) > 0$, then its theta-series is a modular form of weight $1=\frac{2}{2}$ on the subgroup

$$\Gamma_1(N) = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) : a \equiv d \equiv 1 \pmod N, c \equiv 0 \pmod N\}.$$

3) On the other hand, if r is odd, then ϑ_A is a modular form of so-called $\frac{1}{2}$-integral weight $r/2$; such modular forms are defined and discussed at length in [Ko], ch. IV. For example, by taking $Q(x)=x^2$ we obtain the ϑ-series

$$\Theta(z) = \sum_{n=-\infty}^{\infty} e^{2\pi i n^2 z} = \sum_{n\in\mathbb{Z}} q^{n^2}$$

which has weight $\frac{1}{2}$. Note that $\Theta(z)$ is related to the classical theta-function $\theta(z) = \sum_{n\in\mathbb{Z}} e^{-\pi n^2 z}$ of Riemann by the formula $\Theta(z) = \theta(-2iz)$.

1.3 The Space of Modular Forms

1.3.1 Structure theorems

We now turn to study the structure of the spaces $\mathbf{A}$, $\mathbf{M}$ and $\mathbf{S}$ of modular functions, forms and cusp forms. The main result is the following.

Theorem 1.1 *The ring* $\mathbf{M}$ *of all modular forms is the (graded) polynomial ring generated by* E_4 *and* E_6*, and the ideal* $\mathbf{S}$ *of cusp forms is generated by* Δ*:*

$$\mathbf{M} = \mathbb{C}[E_4, E_6] \quad \textit{and} \quad \mathbf{S} = \Delta\mathbf{M}. \tag{1.31}$$

In other words, for every $k \in \mathbb{Z}$*, the set*

$$\mathcal{M}_k := \{E_4^\alpha E_6^\beta : 4\alpha + 6\beta = k, \alpha \geq 0, \beta \geq 0\}$$

is a $\mathbb{C}$*-basis of the space* $\mathbf{M}_k$*, and* $\Delta\mathcal{M}_{k-12}$ *is a* $\mathbb{C}$*-basis of the space* $\mathbf{S}_k = \Delta\mathbf{M}_{k-12}$*. In particular, if* $k < 0$ *or* k *odd, then we have* $\mathbf{M}_k = \mathbf{S}_k = \{0\}$ *and if* k *is even and non-negative, then*

$$\dim \mathbf{M}_k = \begin{cases} \left[\frac{k}{12}\right] & \text{if } k \equiv 2 \ (mod\ 12), \\ \left[\frac{k}{12}\right] + 1 & \text{if } k \not\equiv 2 \ (mod\ 12); \end{cases} \tag{1.32}$$

$$\dim \mathbf{S}_k = \begin{cases} \left[\frac{k}{12}\right] - 1 & \text{if } k \equiv 2 \ (mod\ 12), k \neq 2, \\ \left[\frac{k}{12}\right] & \text{if } k \not\equiv 2 \ (mod\ 12) \text{ or } k = 2. \end{cases} \tag{1.33}$$

Remarks. 1) From the above theorem we thus see that for $k \geq 0$ and k even we have

$$\dim \mathbf{M}_k = 1 \quad \Leftrightarrow \quad k = 0, 4, 6, 8, 10, 14; \tag{1.34}$$

$$\dim \mathbf{S}_k = 0 \quad \Leftrightarrow \quad k \leq 10, \text{ or } k = 14; \tag{1.35}$$

2) Since $\dim M_k < \infty$, it follows that each $f \in M_k$ is determined by a finite set of data. This can in fact be expressed more succinctly in terms of the q-expansion of f, as we shall see in Corollary 1.4 below.

The structure of the field $\mathbf{A}$ of modular functions is given by the following result.

Theorem 1.2 *If* k *is an even integer, then* $\mathbf{A}_k$ *is a one-dimensional* $\mathbf{A}_0$*-vector space generated by* $(E_6/E_4)^{k/2}$ *(whereas* $\mathbf{A}_k = \{0\}$ *if* k *is odd). Furthermore, every modular function of weight* 0 *is a rational function in* j*, i.e.*

$$\mathbf{A}_0 = \mathbb{C}(j) \tag{1.36}$$

is the rational function field generated by the j*-function. In particular,* $\mathbf{A} = \mathbb{C}(E_4, E_6)$ *is the quotient field of* $\mathbf{M}$*.*

1.3.2 Proof of the structure theorems

The main ingredient of the proof of the structure theorems is the following Proposition 1.3, for which we first introduce the following notation.

Notation. Let $f \in \mathcal{M}(\mathfrak{H})$ be a (non-zero) meromorphic function on the upper half plane $\mathfrak{H}$. Then, by definition, f has for each $z_0 \in \mathfrak{H}$ a *Laurent expansion*

$$f(z) = \sum_{n=n_0}^{\infty} a_{n,z_0}(z - z_0)^n \quad \text{in a neighbourhood of } z_0.$$

If $n_0 \in \mathbb{Z}$ has been chosen such that $a_{n_0,z_0} \neq 0$ (as we can always do), then n_0 is called the *order* of the zero or pole of f at z_0 and we write

$$v_{z_0}(f) = \mathrm{ord}_{z_0}(f) = n_0.$$

Note that f is *holomorphic* in a neighbourhood of z_0 if and only if $v_{z_0}(f) \geq 0$.

Similarly, if f has a Fourier expansion of the form

$$f(z) = \sum_{n=n_0}^{\infty} a_n(f) q^n, \quad \text{where } q = e^{2\pi i z},$$

and if $n_0 \in \mathbb{Z}$ has been chosen such that $a_{n_0}(f) \neq 0$, then we call n_0 the *order* of the zero or pole of f at ∞ and write

$$v_\infty(f) = \mathrm{ord}_\infty(f) = n_0.$$

Remark. For any two (non-zero) meromorphic functions $f, g \in \mathcal{M}(\mathfrak{H})$ on $\mathfrak{H}$ and any $z \in \mathfrak{H}$ we have

$$v_z(fg) = v_z(f) + v_z(g) \quad \text{and} \quad v_z(f/g) = v_z(f) - v_z(g). \tag{1.37}$$

Moreover, the same formulae hold for $z = \infty$ provided that f and g are meromorphic at ∞.

We are now ready to state and prove the following *key technical fact* about modular functions on $\mathrm{SL}_2(\mathbb{Z})$:

Proposition 1.3 *If $f \in \mathbf{A}_k$ is a non-zero modular function of weight k, then*

$$v_\infty(f) + \frac{1}{2} v_i(f) + \frac{1}{3} v_\rho(f) + {\sum_{z \in \Gamma\backslash\mathfrak{H}}}^{*} v_z(f) = \frac{k}{12}, \tag{1.38}$$

where the sum run over any system of representatives of $\Gamma\backslash\mathfrak{H}$ which are not Γ-equivalent to i or to $\rho := e^{2\pi i/3}$.

Proof (Sketch). First note that the sum does not depend on the choice of the system of representatives of $\Gamma\backslash\mathfrak{H}$ (i.e. $v_z(f) = v_{\gamma(z)}(f), \forall\gamma \in \Gamma, z \in \mathfrak{H}$); this follows easily from the transformation law of f. Thus, we can choose $\Gamma\backslash\mathfrak{H} \subset \overline{D} = D \cup \partial D$, where

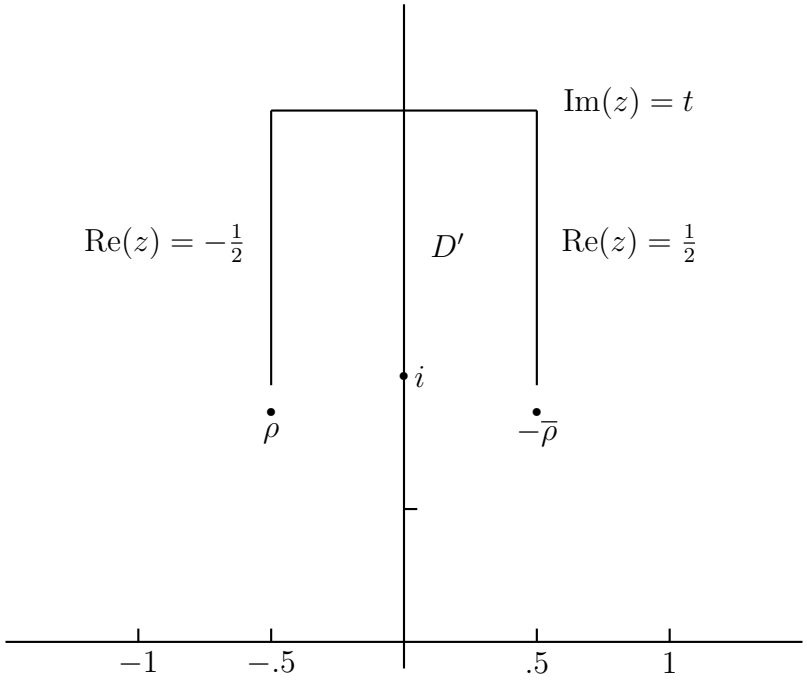
$$D = \{z \in \mathfrak{H} : |z| > 1, |\mathrm{Re}(z)| < \tfrac{1}{2}\}$$

is the so-called *fundamental domain* of $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ (and ∂D is its boundary). (Explicitly, we could take

$$\Gamma\backslash\mathfrak{H} \;=\; D \cup \{z \in \mathfrak{H} : \mathrm{Re}(z) = -\tfrac{1}{2}, |z| \geq 1\} \cup \{z \in \mathfrak{H} : |z| = 1, -\tfrac{1}{2} \leq \mathrm{Re}(z) \leq 0\};$$

cf. [Sch], p. 17.)

We next observe that sum in (1.38) is finite, i.e. that f has only finitely many zeros and poles in D. Indeed, the map $z \mapsto q = e^{2\pi i z}$ defines an isomorphism of D with a subregion of $U^* := \{z \in \mathbb{C} : 0 < |z| < 1\}$. Now since $\tilde{f}(q) = f(z)$ extends to a meromorphic function at $q = 0$ (i.e. at $z = \infty$), $\tilde{f}$ can have only finitely many zeros and poles in U^* and hence the same is true for f in D.



Now fix $r < 1$ and $t > 1$ and consider the region

$$D' = D'(r,t) = \overline{D} \setminus (B(\rho, r) \cup B(-\overline{\rho}, r) \cup B(i, r) \cup \mathfrak{H}(t))$$

in which $B(z_0, r) = \{z \in \mathbb{C} : |z - z_0| < r\}$ and $\mathfrak{H}(t) = \{z \in \mathbb{C} : \mathrm{Im}(z) > t\}$. Then by the residue theorem we have

$$\frac{1}{2\pi i}\int_{\partial D'} \frac{f'}{f} dz = \sum_{z \in D'} v_z(f) = \sum_{z \in \Gamma\backslash\mathfrak{H}}{}^{*} v_z(f),$$

provided that r is sufficiently small and t is sufficiently large (and that f has no zeros or poles on $\partial D'$). On the other hand, by calculating the integral along each piece on the boundary of D' and using the fact that T and S interchange certain pieces of the boundary, we get (by using the transformation law of f under T and S) that

$$\lim_{r\to 0} \frac{1}{2\pi i} \int_{\partial D'(r,t)} \frac{f'}{f} dz \;=\; -\tfrac{1}{2} v_i(f) - \tfrac{1}{3} v_\rho(f) + \tfrac{k}{12};$$

cf. [Se1], p. 86 or [Ko], p. 116 for the precise details. This proves (1.38) in the case that f has no zeros or poles on $\partial D'$. For the general case the proof is similar, except that the region D' is more complicated: one also removes (and adds) small disks around the zeros and poles of f which lie on ∂D.

Corollary 1.4 *Two modular forms f_1 and $f_2 \in \mathbf{M}_k$ of weight k are identical if and only if their Fourier coefficients $a_n(f_i)$ coincide for $n \leq \left[\frac{k}{12}\right]$; i.e.*

$$a_n(f_1) = a_n(f_2), \text{ for all } n \leq \left[\tfrac{k}{12}\right] \quad \Rightarrow \quad f_1 = f_2.$$

Proof. Put $f = f_1 - f_2 \in \mathbf{M}_k$. Then by hypothesis $v_\infty(f) \geq \left[\frac{k}{12}\right] + 1 > \frac{k}{12}$. Since $v_z(f) \geq 0, \forall z \in \mathfrak{H}$, this contradicts (1.38) and so f must be zero, i.e. $f_1 = f_2$.

Before proving Theorem 1.1, we first prove the following results which are in fact special cases of Theorem 1.1.

Proposition 1.5 *(a)* $\mathbf{M}_0 = \mathbb{C} \cdot 1$.

(b) $\mathbf{M}_k = \{0\}$ *if* $k < 0$, $k = 2$ *or if* k *is odd.*

(c) $\mathbf{M}_k = \mathbb{C}E_k$ *for* $k = 4$, 6, 8, 10 *or* 14.

(d) $\mathbf{S}_k = \{0\}$ *for* $k < 12$ *and* $\mathbf{S}_{12} = \mathbb{C}\Delta$.

(e) $\mathbf{S}_k = \Delta \mathbf{M}_{k-12}$, *for all* $k \in \mathbb{Z}$.

(f) $\mathbf{M}_k = \mathbf{S}_k \oplus \mathbb{C}E_k$, *for all* $k \geq 4$.

Proof. First note that if $f \in \mathbf{M}_k$, then f is holomorphic everywhere and so $v_z(f) \geq 0$, for all $z \in \mathfrak{H} \cup \{\infty\}$.

(a) Fix $z_0 \in \mathfrak{H}$. If $f \in \mathbf{M}_0$, then also $f_1 := f - f(z_0) \cdot 1 \in \mathbf{M}_0$. But $v_{z_0}(f_1) > 0$ by construction, and this contradicts (1.38) unless $f_1 = 0$. Thus $f = f(z_0) \cdot 1$ is constant, so $\mathbf{M}_0 = \mathbb{C} \cdot 1$.

(b) If $k < 0$, then for any non-zero $f \in \mathbf{M}_k$ the left hand side of (1.38) is non-negative, whereas the right hand side is negative. Thus $\mathbf{M}_k = \{0\}$.

Similarly, if $k = 2$, then the right hand side of (1.38) is $\frac{1}{6}$ whereas the left hand side is either 0 or $\geq \frac{1}{3}$, and so $\mathbf{M}_2 = \{0\}$.

Finally, the fact that $\mathbf{M}_k = \{0\}$ if k is odd was already mentioned earlier (cf. §1.1).

(c) If $k = 4, 6, 8, 10$, or 14, then (1.38) has only one possible solution:

$$\begin{array}{lll} k=4: & v_\rho(f)=1, & v_z(f)=0, \forall z \neq \rho \\ k=6: & v_i(f)=1, & v_z(f)=0, \forall z \neq i \\ k=8: & v_\rho(f)=2, & v_z(f)=0, \forall z \neq \rho \\ k=10: & v_i(f)=v_\rho(f)=1, & v_z(f)=0, \forall z \neq i, \rho \\ k=14: & v_i(f)=1, v_\rho(f)=2, & v_z(f)=0, \forall z \neq i, \rho \end{array}$$

Now let $f_1, f_2 \in \mathbf{M}_k$ ($f_i \neq 0$). Then by the above we know that f_1 and f_2 have the same orders of zeros, and hence f_1/f_2 is holomorphic on $\mathfrak{H} \cup \{\infty\}$. Thus $f_1/f_2 \in \mathbf{M}_0$, and so by part (a) we have $f_1 = cf_2$, for some $c \in \mathbb{C}$. Taking $f_2 = E_k \in \mathbf{M}_k$ yields the assertion.

(d) If $f \in \mathbf{S}_k$ is a cusp form, then $v_\infty(f) \geq 1$ (by definition). Since the right hand side of (1.38) is < 1 for $k < 12$, it follows that $\mathbf{S}_k = \{0\}$.

Moreover, for $k = 12$ we see from (1.38) that every non-zero $f \in \mathbf{S}_{12}$ satisfies

$$v_\infty(f) = 1 \quad \text{and} \quad v_z(f) = 0, \ \forall z \in \mathfrak{H}; \tag{1.39}$$

in particular this holds for $f = \Delta$. Thus, by the same argument as in (c) we see that $\mathbf{S}_{12} = \mathbb{C}\Delta$.

(e) If $f \in \mathbf{S}_k$ is a (non-zero) cusp form, then $v_\infty(f/\Delta) \geq 1 - 1 = 0$ and $v_z(f/\Delta) = v_z(f), \forall z \in \mathfrak{H}$, by (1.39), and so $f/\Delta \in \mathbf{M}_{k-12}$. Thus $\mathbf{S}_k \subset \Delta\mathbf{M}_{k-12}$, and so we have the desired equality since the opposite inclusion is obvious.

(f) Clearly $cE_k \notin \mathbf{S}_k$, if $c \neq 0$, so $\mathbf{S}_k \cap \mathbb{C}E_k = \{0\}$. Moreover, if $f \in \mathbf{M}_k$, then $f - a_0(f)E_k \in \mathbf{S}_k$ (because $a_0(E_k) = 1$), and so the assertion follows.

Remark. For later reference, note that in part (c) of the above proof we had shown:

$$v_\rho(E_4) = 1 \quad \text{and} \quad v_z(E_4) = 0, \text{ for all } z \neq \rho \tag{1.40}$$

$$v_i(E_6) = 1 \quad \text{and} \quad v_z(E_6) = 0, \text{ for all } z \neq i. \tag{1.41}$$

Proof of Theorem 1.1. By Proposition 1.5(d) we know that $\mathbf{S}_k = \Delta\mathbf{M}_{k-12}, \forall k \in \mathbb{Z}$ and so $\mathbf{S} = \Delta\mathbf{M}$. It thus remains to show that $\mathbf{M} = \mathbb{C}[E_4, E_6]$ or equivalently, that $\mathcal{M}_k$ is a basis of M_k.

Claim 1. $\mathcal{M}_k$ generates $\mathbf{M}_k$, i.e. $\mathbf{M}_k = \langle \mathcal{M}_k \rangle$, for all $k \in \mathbb{Z}$.

This is clear if k is odd, so assume k even. By Proposition 1.5(a)-(c), this is trivial for $k \leq 2$ (or for k odd). To prove that it is true in general, induct on $k \geq 4$ (and assume that k is even). For $k \geq 4$ it is immediate that $\mathcal{M}_k \neq \emptyset$, so let $f_k \in \mathcal{M}_k$. If $f \in \mathbf{M}_k$, then $g := f - a_0(f)f_k \in \mathbf{S}_k$ because $a_0(f_k) = 1$, and so by Proposition 1.5(e) we have $g = \Delta h$ with $h \in \mathbf{M}_{k-12}$. By induction, $\mathbf{M}_{k-12} = \langle \mathcal{M}_{k-12} \rangle$ so $f \in \langle f_k, \Delta\mathcal{M}_{k-12} \rangle$. Now since $\Delta = \frac{(2\pi)^{12}}{12^3}(E_4^3 - E_6^2)$, it follows that $\Delta\mathcal{M}_{k-12} \subset \langle \mathcal{M}_k \rangle$, and so $f \in \langle \mathcal{M}_k \rangle$. This proves the inclusion $\mathbf{M}_k \subset \langle \mathcal{M}_k \rangle$, and so we have the desired equality since the opposite inclusion is trivial.

Claim 2. If $k \geq 0$ is even, then $\#\mathcal{M}_k = \begin{cases} \left[\frac{k}{12}\right] & \text{if } k \equiv 2 \pmod{12}, \\ \left[\frac{k}{12}\right] + 1 & \text{if } k \not\equiv 2 \pmod{12}. \end{cases}$

By definition, $\#\mathcal{M}_k$ is the number of non-negative integer solutions (α, β) of the linear Diophantine equation $4\alpha + 6\beta = k$. Since the general integer solution of this equation is $\alpha = -\frac{k}{2} + 3t$, $\beta = \frac{k}{2} - 2t$, where $t \in \mathbb{Z}$, it follows that

$$\#\mathcal{M}_k = \#\{t \in \mathbb{Z} : \tfrac{k}{6} \le t \le \tfrac{k}{4}\} = \begin{cases} \left[\frac{k}{4}\right] - \left[\frac{k}{6}\right] + 1 & \text{if } \frac{k}{6} \in \mathbb{Z}, \\ \left[\frac{k}{4}\right] - \left[\frac{k}{6}\right] & \text{otherwise.} \end{cases}$$

Thus, if $k \equiv 0, 6 \pmod{12}$, the assertion of Claim 2 follows. For the other cases write $k = 12k' + r$, with $0 \le r < 12$. Then $[\frac{k}{4}] - [\frac{k}{6}] = k' + [\frac{r}{4}] - [\frac{r}{6}]$. Now since $[\frac{r}{4}] - [\frac{r}{6}] = 1$ (resp. $= 0$) if $r = 4, 8, 10$ (resp. if $r = 2$), the assertion follows in the other cases as well.

Claim 3. $\dim \mathbf{M}_k = \#\mathcal{M}_k$, for all k.

Again we induct on k (where k is even). For $k < 12$ this is clear by Proposition 1.5(c). For $k \ge 12$ we have $\dim \mathbf{M}_k = 1 + \dim \mathbf{S}_k = 1 + \dim \mathbf{M}_{k-12}$ by Proposition 1.5(e),(f). On the other hand, by Claim 2 we see that $\#\mathcal{M}_k = 1 + \#\mathcal{M}_{k-12}$, for all $k \ge 12$ and so Claim 3 follows.

By Claims 1 and 3 we thus see that $\mathcal{M}_k$ is a basis of $\mathbf{M}_k$. Moreover, formula (1.32) follows from Claim 2 and formula (1.33) follows from this and the fact that $\dim \mathbf{S}_k = \dim \mathbf{M}_k - 1$ for $k \ge 4$.

Proof of Theorem 1.2. The first assertion is easily verified. Indeed, let $f_k := (E_6/E_4)^{k/2}$. Then $0 \ne f_k \in \mathbf{A}_k$, and so we see that for any modular function $f \in \mathbf{A}_k$ of weight k, the quotient $f/f_k \in \mathbf{A}_0$ is a modular function of weight 0, which means that $\mathbf{A}_k = \mathbf{A}_0 f_k$, as claimed.

To prove the second assertion, i.e. that $\mathbf{A}_0 = \mathbb{C}(j)$, recall first that $j = \frac{E_4^3}{\Delta_1}$, where $\Delta_1 = \frac{1}{12^3}(E_4^3 - E_6^2)$; cf. equation (1.17). Then $j - 12^3 = \frac{E_6^2}{\Delta_1}$ and so

$$j^\alpha (j - 12^3)^\beta = \frac{E_4^{3\alpha} E_6^{2\beta}}{\Delta_1^{\alpha+\beta}}, \quad \text{for all } \alpha, \beta \ge 0. \tag{1.42}$$

Now suppose that $f \in \mathbf{A}_0$ is holomorphic on $\mathfrak{H}$ and that $f \ne 0$. Then by (1.38) we see that $-\nu := v_\infty(f) \le 0$, and so by (1.39) we see that $g := f\Delta^\nu \in \mathbf{M}_{12\nu}$. Thus, by Theorem 1.1 we know that g is a linear combination of terms of the form $h := E_4^\alpha E_6^\beta$ with $4\alpha + 6\beta = 12\nu$. Thus $\alpha = 3\alpha'$ and $\beta = 2\beta'$, and so h/Δ_1^ν has the form of the right side of (1.42), which means that $h/\Delta_1^\nu \in \mathbb{C}[j]$. Thus also $f = g/\Delta_1^\nu \in \mathbb{C}[j]$, and so we have shown:

$$f \in \mathbf{A}_0,\ f \text{ holomorphic on } \mathfrak{H} \quad \Rightarrow \quad f \in \mathbb{C}[j] \tag{1.43}$$

Now suppose that $f \in \mathbf{A}_0$ is arbitrary, and let $z_1, \dots, z_r$ denote the poles of f on $\Gamma\backslash\mathfrak{H}$ (i.e. in $\overline{D}$), and $n_1, \dots, n_r$ their corresponding multiplicities. Then

$$f_1 = f \prod (j(z) - j(z_i))^{n_i} \in \mathbf{A}_0$$

is holomorphic on $\mathfrak{H}$ and so $f_1 \in \mathbb{C}[j]$ by (1.43). Thus $f \in \mathbb{C}(j)$ (which is the quotient field of $\mathbb{C}[j]$).

1.3.3 Application 1: Identities between arithmetical functions

As a first application, we shall see that the theory developed so far suffices to prove a number of interesting identities for the arithmetical functions $\sigma_k(n)$. In all cases, these identities are just a translation of the corresponding identities among products of the E_k's such as the following.

Proposition 1.6 *We have $E_4^2 = E_8$, $E_4E_6 = E_{10}$, and $E_6E_8 = E_4E_{10} = E_{14}$.*

Proof. Clearly $E_4^2, E_8 \in \mathbf{M}_8$. By (1.34) we have $\dim \mathbf{M}_8 = 1$, so $E_4^2 = cE_8$, for some $c \in \mathbb{C}$. Since the q-expansions of both E_4^2 and E_8 have constant term 1, we have $c = 1$, or $E_4^2 = E_8$. The other identities are proved similarly.

Corollary 1.7 *The following identities hold:*

$$\begin{aligned} 120\sum_{k=1}^{n-1}\sigma_3(k)\sigma_3(n-k) &= \sigma_7(n) - \sigma_3(n) \\ 5040\sum_{k=1}^{n-1}\sigma_3(k)\sigma_5(n-k) &= 11\sigma_9(n) - 21\sigma_5(n) + 10\sigma_3(n) \\ 10080\sum_{k=1}^{n-1}\sigma_5(k)\sigma_7(n-k) &= \sigma_{13}(n) + 20\sigma_7(n) - 21\sigma_5(n). \end{aligned}$$

Proof. For any even integers $r, s \geq 2$ we have by (1.8) that

$$E_rE_s = 1 + c_rc_s\sum_{n=1}^{\infty}\left(\frac{\sigma_{r-1}(n)}{c_s} + \frac{\sigma_{s-1}(n)}{c_r} + \sum_{m=1}^{n-1}\sigma_{r-1}(m)\sigma_{s-1}(n-m)\right)q^n. \tag{1.44}$$

Thus, the identities given in the corollary are just restatements of the identities $E_4^2 = E_8$, $E_4E_6 = E_{10}$, and $E_6E_8 = E_{14}$, respectively.

Proposition 1.8 *We have $E_{12} - E_6^2 = 12^3\frac{441}{691}\Delta_1 = \frac{441}{691}(E_4^3 - E_6^2)$.*

Proof. The functions $E_{12} - E_6^2$ and Δ_1 are both cusp forms of weight 12, so $E_{12} - E_6^2 = c\Delta_1$ for some $c \in \mathbb{C}$ because $\dim \mathbf{S}_{12} = 1$; cf. (1.33). To determine c, we look at the q-expansions of both functions. Since $a_1(\Delta_1) = 1 \neq 0$, we see that $c = a_1(E_{12} - E_6^2)/a_1(\Delta_1) = (\frac{65520}{691} - (-1008)) = \frac{762048}{691} = \frac{1728\cdot 441}{691}$, as claimed.

Corollary 1.9 *The following identity holds:*

$$\tau(n) = \frac{65}{756}\sigma_{11}(n) - \frac{691}{756}\sigma_5(n) + \frac{691}{3}\sum_{k=1}^{n-1}\sigma_5(k)\sigma_5(n-k)$$

In particular, we have the congruence

$$\tau(n) \equiv \sigma_{11}(n) \ (mod\ 691).$$

Proof. Since $\Delta_1 = \sum_{n\geq 1}\tau(n)q^n$, the first identity is just a restatement of Proposition 1.8, and from this the given congruence follows because $65 \equiv 756 \pmod{691}$ (and 691 is prime.)

Although E_2 isn't a modular form, it still gives rise to interesting identities:

Proposition 1.10 (Ramanujan[Ra]) *Let $h_k = \theta E_k - \frac{k}{12}(E_2E_k - E_{k+2})$, where $k \geq 2$ and θ is the derivative operator; cf. (1.13). If $k \geq 4$, then $h_k \in \mathbf{S}_{k+2}$, whereas $h_2 = -\theta E_2$. In particular, we have the following identities:*

$$\begin{aligned} \theta E_2 &= \tfrac{1}{12}(E_2^2 - E_4), & \theta E_4 &= \tfrac{1}{3}(E_2E_4 - E_6), \\ \theta E_6 &= \tfrac{1}{2}(E_2E_6 - E_8), & \theta E_8 &= \tfrac{2}{3}(E_2E_6 - E_{10}). \end{aligned}$$

Proof. Since $E_k \in \mathbf{M}_k$ for $k \geq 4$, the first assertion follows immediately by applying Ramanujan's observation (1.12) to $f = E_k$. Moreover, since $\mathbf{S}_{k+2} = 0$ for $k \leq 8$, we see that $h_4 = h_6 = h_8 = 0$, and this yields the last three displayed identities. Now if $k = 2$, then by differentiating the functional equation (1.11) of E_2 we obtain that $\theta E_2 - \frac{1}{12}E_2^2 \in \mathbf{M}_4 = \mathbb{C}E_4$, and so we see that $\theta E_2 = \frac{1}{12}(E_2^2 - E_4)$ (by looking at the q-expansions). Thus $h_2 = \theta E_2 - \frac{2}{12}(E_2^2 - E_4) = \theta E_2 - 2\theta E_2 = -\theta E_2$, as claimed.

As before, these identities are equivalent to certain identities involving the sigma functions σ_k; the first of these was discovered by Glaisher[Gl] in 1884:

Corollary 1.11 *The following identities hold for all $n \geq 1$:*

$$\begin{aligned} \sum_{k=1}^{n-1}\sigma(k)\sigma(n-k) &= \frac{1}{12}\left[5\sigma_3(n) - (6n-1)\sigma(n)\right], \\ \sum_{k=1}^{n-1}\sigma(k)\sigma_3(n-k) &= \frac{1}{240}\left[21\sigma_5(n) - 10(3n-1)\sigma_3(n) - \sigma(n)\right], \\ \sum_{k=1}^{n-1}\sigma(k)\sigma_5(n-k) &= \frac{1}{504}\left[20\sigma_7(n) - 21(2n-1)\sigma_5(n) + \sigma(n)\right], \\ \sum_{k=1}^{n-1}\sigma(k)\sigma_7(n-k) &= \frac{1}{480}\left[11\sigma_9(n) - 10(3n-2)\sigma_7(n) - \sigma(n)\right]. \end{aligned}$$

Corollary 1.12 *The derivative operator θ maps the ring $\widetilde{\mathbf{M}} := \mathbb{C}[E_2, E_4, E_6]$ of "quasi-modular forms" into itself.*

Proof. By the product rule of derivatives, it is enough to verify that $\theta E_2, \theta E_4, \theta E_6 \in \widetilde{\mathbf{M}}$, and this is clear by the identities of Proposition 1.10.

Remark. The basic theory of *quasi-modular forms* is presented in Kaneko/Zagier[KZ], and connections of this theory to *String Theory* and to *Mirror Symmetry* in Physics are explained in Dijkgraaf's article[Dij]. See also Lang[La], p. 161.

1.3.4 Estimates for the Fourier coefficients of modular forms

We next want to study the growth rate of the Fourier coefficients $a_n = a_n(f)$ of a modular form

$$f(z) = \sum_{n=0}^{\infty} a_n q^n \quad (\text{where } q = e^{2\pi i z}).$$

For $f = E_k$, where $k \geq 4$ is even, we have the estimate

$$a_n(f) = O(n^{k-1}). \tag{1.45}$$

This follows easily from the following more precise assertion which also shows that the above estimate is best possible.

Proposition 1.13 *We have the estimates*

$$\begin{aligned} n &< \sigma(n) < n(1+\log(n)), \\ n^k &< \sigma_k(n) < \zeta(k)n^k, \text{ if } k > 1. \end{aligned}$$

Proof. This is well-known; cf. [Sch], p. 224. Indeed, we have

$$n < \sigma(n) = n \sum_{0<d|n} \frac{1}{d} < n \sum_{\nu=1}^{n} \frac{1}{\nu} < n(1+\log(n)).$$

which yields the desired bounds on $\sigma(n)$. Similarly, for $k > 1$

$$n^k < \sigma_k(n) = n^k \sum_{0<d|n} \frac{1}{d^k} < n^k \sum_{\nu=1}^{\infty} \frac{1}{\nu^k} = \zeta(k)n^k.$$

Since the space of modular forms is generated by products of E_4 and E_6, we obtain

Corollary 1.14 *The estimate* (1.45) *holds for any modular form* $f \in \mathbf{M}_k$.

On the other hand, if $f \in \mathbf{S}_k$ is a cusp form, then much better estimates are valid.

Theorem 1.15 (Hecke) *If* f *is a cusp form of weight* k*, then*

$$a_n(f) = O(n^{k/2}). \tag{1.46}$$

Remark. The above estimate (1.46) is by no means the best. Indeed, it follows from the *Petersson-Ramanujan Conjecture* which generalizes the Ramanujan Conjecture mentioned in subsection 1.2.2 (and which was also proved by Deligne) that one has the estimate

$$a_n(f) = O(n^{k/2-1/2+\epsilon}), \quad \text{for } f \in \mathbf{S}_k. \tag{1.47}$$

This estimate, however, is in fact best possible. Indeed, Ramanujan[Ra] shows that it follows from his conjecture(s) that one has

$$\tau(n) \geq n^{11/2}$$

for infinitely many n of the form $n = p^r$ (where p is any prime such that $\tau(p) \neq 0$), and the same reasoning extends to prove a similar result for a suitable basis of $\mathbf{S}_k$, for any k.

Proof (of Theorem 1.15). Since $a_0 = 0$ we have $f = q(\sum a_n q^{n-1})$, and so

$$|f(z)| = O(q) = O(e^{-2\pi Im(z)}), \quad \text{as } q \to 0.$$

Next, we observe that the transformation law of f under Γ shows that the function $\phi(z) := |f(z)|\mathrm{Im}(z)^{k/2}$ is invariant under Γ and hence (by the above) is bounded on all of $\mathfrak{H}$. Thus for some constant M we have

$$|f(z)| \leq M(\mathrm{Im}(z))^{-k/2}, \quad \text{for all } z \in \mathfrak{H}.$$

Thus, by using the integral representation

$$a_n = \int_0^1 f(x+iy)q^{-n}dx,$$

of the Fourier coefficients of a periodic function, we get the estimate

$$|a_n| \leq M(\mathrm{Im}(z))^{-\frac{k}{2}} e^{2\pi n Im(z)},$$

which is true for all $z \in \mathfrak{H}$. In particular, if we take z such that $\mathrm{Im}(z) = \frac{1}{n}$, then we obtain the assertion of the theorem.

Corollary 1.16 *For any modular form $f \in \mathbf{M}_k$ of even weight $k \geq 4$ we have*

$$a_n(f) = c_f \sigma_{k-1}(n) + O(n^{k/2}), \quad \textit{where } c_f = c_k a_0(f). \tag{1.48}$$

In particular, if $a_0(f) \neq 0$, then the order of magnitude of $a_n(f)$ is n^{k-1}, i.e. there exist constants c_1, c_2 such that

$$c_1 n^{k-1} < |a_n(f)| < c_2 n^{k-1}. \tag{1.49}$$

Proof. Put $g = f - a_0(f)E_k$. Then $g \in \mathbf{S}_k$, so $a_n(g) = O(n^{k/2})$ by Hecke's Theorem. On the other hand, $a_n(g) = a_n(f) - a_0(f)a_n(E_k) = a_n(f) - a_0(f)a_n(E_k) = a_n(f) - c_f\sigma_{k-1}(n)$, and so (1.48) follows. Furthermore, if $a_n(f) \neq 0$, then also $c_f \neq 0$ and so (1.49) follows from (1.48) and Proposition 1.13.

1.3.5 Application 2: The order of magnitude of arithmetical functions

In application 1 we used the knowledge of explicit modular forms to derive identities among certain arithmetical functions involving the functions σ_k, at least for small values of k. For larger values of k this method becomes infeasible since the expressions become too involved to be interesting or useful. However, since the coefficients of cusp forms have a slower growth rate than the σ_k's, (cf. Corollary 1.16 above), we can combine the uninteresting terms into an error term and thus obtain powerful results on the order of magnitude of certain arithmetic functions. As an example of this method, let us consider the arithmetical functions $\Sigma_{r,s}$ which were studied by Ramanujan in his monumental paper [Ra]. Further examples will appear in the next subsection 1.3.6.

Following Ramanujan, let us put, using the notation of subsection 1.2.1,

$$\sigma_{k-1}(0) \stackrel{\text{def}}{=} -\frac{B_k}{2k} = \frac{1}{c_k},$$

so that we can now write equation (1.8) in the form

$$E_k(z) = c_k \sum_{k=0}^{\infty} \sigma_{k-1}(n) q^n. \tag{1.50}$$

Again following Ramanujan[Ra], let us consider the function

$$\Sigma_{r,s}(n) = \sum_{k=0}^{n} \sigma_r(k)\sigma_s(n-k), \tag{1.51}$$

where r, s are *odd* positive integers. (By symmetry we may assume that $r \leq s$.) From equation (1.50) we see that the generating function of this function is

$$\sum_{n=0}^{\infty} \Sigma_{r,s}(n) q^n = \frac{1}{c_{r+1}c_{s+1}} E_{r+1}(z)E_{s+1}(z) = \frac{1}{4}\zeta(-r)\zeta(-s)E_{r+1}(z)E_{s+1}(z),$$

where (in the second equality) we have used the identity $\zeta(-r) = \frac{2}{c_{r+1}}$ which follows from Euler's formula for $\zeta(r+1)$ and the functional equation (1.54) below.

Following Ramanujan, the growth rate of $\Sigma_{r,s}$ can be expressed as follows.

Theorem 1.17 (Ramanujan) *If r and s are odd positive integers, then we have*

$$\Sigma_{r,s}(n) = \frac{\zeta(-r)\zeta(-s)}{2\zeta(-r-s-1)}\sigma_{r+s+1}(n) + \frac{\zeta(1-r)+\zeta(1-s)}{r+s} n\sigma_{r+s-1}(n) + O(n^{\frac{1}{2}(r+s)+1}). \tag{1.52}$$

Furthermore, in the 9 cases that $r+s \leq 12$ and $r+s \neq 10$ there is no error term in (1.52); i.e. we have in these cases the identities

$$\Sigma_{r,s}(n) = \frac{\zeta(-r)\zeta(-s)}{2\zeta(-r-s-1)}\sigma_{r+s+1}(n) + \frac{\zeta(1-r)+\zeta(1-s)}{r+s} n\sigma_{r+s-1}(n). \tag{1.53}$$

Proof. Suppose first that $r > 1$ and $s > 1$. Then $f = E_{r+1}E_{s+1} - E_{r+s+2}$ is a cusp form of weight $t = r + s + 2$, so by Hecke's theorem (1.15) we obtain

$$\Sigma r, s(n)/(\zeta(-r)\zeta(-s)) - \sigma_{r+s+1}(n)/(2\zeta(-r-s-1)) = O(n^{\frac{1}{2}(r+s+2)}).$$

Since in this case $\zeta(1-r) = \zeta(1-s) = 0$, this equation is equivalent to equation (1.52). Furthermore, from (1.35) we know that $\mathbf{S}_t = 0$ for $t \leq 14$, $t \neq 12$, and so the identity (1.53) holds.

Next, suppose that $r = 1$. If $s > 1$, then by Proposition 1.10 we know that $h_{s+1} = \theta E_{s+1} - \frac{s+1}{12}(E_2E_{s+1} - E_{s+3}) \in \mathbf{S}_{s+3}$ is a cusp form of weight $s + 3$. From this, the estimate (1.52) follows readily from Hecke's theorem (1.15) since $\zeta(0) = -\frac{1}{2}$. Finally, if $r = s = 1$, then (1.53) is just a restatement of Glaisher's identity (i.e. the first identity of Corollary 1.11).

Remarks. 1) Note that the identities of (1.53) constitute a succinct way of writing the 7 explicit identities of Corollaries 1.7 and 1.11. In fact, (1.53) also includes two more identities which were not mentioned earlier:

$$\begin{aligned}
\sum_{k=1}^{n-1} \sigma_3(k)\sigma_9(n-k) &= \frac{1}{2640}\left[\sigma_{13}(n) - 11\sigma_9(n) + 10\sigma_3(n)\right] \\
\sum_{k=1}^{n-1} \sigma(k)\sigma_{11}(n-k) &= \frac{1}{65520}\left[691\sigma_{13}(n) - 2730(n-1)\sigma_{11}(n) - 691\sigma(n)\right].
\end{aligned}$$

2) Ramanujan[Ra] did not have Hecke's theorem available, so he could only prove (by using an "elementary" method) that the above error term is $O(n^{\frac{2}{3}(r+s+1)})$. However, in the same paper he conjectured that the error term is in fact $O(n^{\frac{1}{2}(r+s+1+\epsilon)})$ and showed that it cannot be smaller than $O(n^{\frac{1}{2}(r+s+1)})$. It follows again by the Petersson-Ramanujan Conjecture (proved by Deligne) that this (best) error estimate is indeed correct.

3) In Ramanujan's paper [Ra] the coefficient of $\sigma_{r+s+1}(n)$ in formula (1.52) is given as

$$\frac{\Gamma(r+1)\Gamma(s+1)}{\Gamma(r+s+2)}\frac{\zeta(r+1)\zeta(s+1)}{\zeta(r+s+2)}.$$

This is in fact equal to the coefficient given above because from the functional equation of the ζ-function,

$$\zeta(1-s) = 2^{1-s}\pi^{-s}\cos\left(\frac{\pi s}{2}\right)\Gamma(s)\zeta(s), \tag{1.54}$$

we obtain, if r is an odd integer,

$$\Gamma(r+1)\zeta(r+1) = (-1)^{\frac{r+1}{2}}2^r\pi^{r+1}\zeta(-r),$$

and from this (and Euler's formula) the relations

$$\frac{\Gamma(r+1)\Gamma(s+1)}{\Gamma(r+s+2)}\frac{\zeta(r+1)\zeta(s+1)}{\zeta(r+s+2)} = \frac{\zeta(-r)\zeta(-s)}{2\zeta(-r-s-1)} = \frac{c_{r+s+1}}{c_{r+1}c_{s+1}}$$

follow readily.

1.3.6 Application 3: Unimodular lattices

As yet another application, let us consider even integral lattices $L \subset \mathbb{R}^r$ of dimension r. This means:

- $L \simeq \mathbb{Z}^r$ as an abelian group, and L contains a basis of $\mathbb{R}^r$;
- the usual dot product $(\,\cdot\,)$ on $\mathbb{R}^r$ assumes integral values on $L \times L$ and even integral values on the diagonal of $L \times L$.

A basic question in the theory of lattices is to calculate or to estimate the number

$$r_L(n) = \#\{\mathbf{x} \in L : (\mathbf{x} \cdot \mathbf{x}) = n\}$$

of lattice vectors of a given squared-length n.

By fixing a basis of L and hence an isomorphism $L \simeq \mathbb{Z}^r$, we can equivalently think of a lattice as the module $\mathbb{Z}^r$ endowed with an even integral positive-definite quadratic form $Q(x_1, \ldots, x_r)$. Explicitly, if $\mathbf{v} := \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is a basis of L, then $Q = Q_{L,\mathbf{v}}$ is given by

$$Q(x_1, x_2, \ldots, x_r) = \tfrac{1}{2}||x_1\mathbf{v}_1 + \ldots x_r\mathbf{v}_r||^2 = \tfrac{1}{2}\vec{x}^t A\vec{x}, \quad \text{where } A = ((\mathbf{v}_i \cdot \mathbf{v}_j))_{i,j}.$$

In this translation, the number $r_L(2n)$ of lattice-vectors of squared-length $2n$ becomes the number $r_Q(n)$ of representations of n by Q, i.e.,

$$r_L(2n) = r_Q(n).$$

Thus, an equivalent formulation of the above problem is to calculate or to estimate the number of representations of a number n by an even integral positive-definite quadratic form Q.

Let us now assume in addition that L (or Q) is *unimodular*, i.e. that its volume $\mathrm{vol}(L) := \det(A) = 1$. Then, by the discussion in subsection (1.2.5), we know that the associated theta-series

$$\vartheta_L(z) = \vartheta_Q(z) = \sum_{n=0}^{\infty} r_L(2n)q^n$$

defines a modular form of weight $r/2$; recall that we necessarily have that $r \equiv 0 \pmod 8$. Since $r_L(0) = 1$, we see that

$$f_L = \vartheta_L - E_{r/2} \in \mathbf{S}_{r/2} \tag{1.55}$$

is a cusp form of weight $r/2$. Thus, by Theorem 1.15 we obtain:

Theorem 1.18 *If L is an even unimodular lattice of dimension r, then the number $r_L(2n)$ of lattice-vectors of squared-length $2n$ satisfies:*

$$r_L(2n) = \frac{r}{B_{\frac{r}{2}}}\sigma_{\frac{r}{2}-1}(n) + O(n^{r/4}). \tag{1.56}$$

Examples 0) For any $r = 4m$, the set

$$\Gamma_r = \{(x_1, \dots x_r) \in \tfrac{1}{2}\mathbb{Z}^r : \textstyle\sum x_i \in 2\mathbb{Z}, x_i - x_j \in \mathbb{Z}, \forall i, j\}$$

defines an integral lattice of determinant 1, as is easy to see; cf. [Se1], p. 51. Furthermore, Γ_r is an *even* integral lattice if (and only if) $r \equiv 0 \pmod 8$. Thus, for any such r, the number $r_{\Gamma_r}(2n)$ satisfies the growth rate (1.56).

1) $r = 8$. If L is an even unimodular lattice of dimension 8, then we have

$$r_L(2n) = 240\sigma_3(n), \quad \text{for all } n \in \mathbb{N},$$

because there are no non-zero cusp forms of weight $r/2 = 4$ (and so $f_L = 0$). It is known, however, that up to isomorphism there is only one such lattice, namely the lattice Γ_8 arising from the exceptional Lie algebra E_8; cf. [CS], p. 423.

2) $r = 16$. Here again there are no non-zero cusp forms of weight $r/2 = 8$, and so

$$r_L(2n) = 480\sigma_7(n), \quad \text{for all } n \in \mathbb{N},$$

for every even unimodular lattice L of dimension 16. In this case there are two such (non-isomorphic) lattices: $\Gamma_8 \oplus \Gamma_8$ and Γ_{16} (in the notation of Example 0)).

3) $r = 24$. If L is an even unimodular lattice of dimension 24, then there is a constant $c_L \in \mathbb{Q}$ such that

$$\vartheta_L = E_{12} + \frac{c_L}{(2\pi)^{12}}\Delta,$$

because $\mathbf{S}_{12} = \mathbb{C}\Delta$ is generated by Δ. This means:

$$r_L(2n) = \frac{65520}{691}\sigma_{11}(n) + c_L\tau(n), \quad \text{for all } n \in \mathbb{N},$$

and so the constant c_L is determined by

$$c_L = r_L(2) - \frac{65520}{691}.$$

By a theorem of Niemeier (1968) it is known that there are exactly 24 non-isomorphic even unimodular lattices L of dimension 24; cf. Conway/Sloane[CS], ch. 16, 18. (For any $r = 8m$, there is a general formula for the weighted number of even unimodular lattices in terms of Bernoulli numbers; cf. [CS], p. 409.) Four of these are the following:

a) $L = \Gamma_{24}$. Here $r_L(2) = 2 \cdot 24 \cdot 23$, so $c_L = \frac{697344}{691}$.

b) $L = \Gamma_8 \oplus \Gamma_8 \oplus \Gamma_8$. Here $r_L(2) = 3r_{\Gamma_8}(2) = 3 \cdot 240$, so $c_L = \frac{432000}{691}$.

c) $L = \Gamma_8 \oplus \Gamma_{16}$. Here $r_L(2) = r_{\Gamma_8}(2) + r_{\Gamma_{16}}(2) = 720$, so again $c_L = \frac{432000}{691}$.

d) $L =$ Leech lattice. This the even unimodular lattice of dimension 24 which is characterized by the condition that $r_L(2) = 0$; thus, in this case $c_L = -\frac{65520}{691}$. We thus see that the shortest non-zero vector in L has squared-length 4, and that there are $r_L(4) = \frac{65520}{691}(\sigma_{11}(2) - \tau(2)) = 196560$ such vectors!

1.4 Modular Interpretation

The term "modular" comes from the Latin word *modulus* = measure, standard of measurement. What is being measured here are elliptic curves: (elliptic) modular functions are functions that measure properties of elliptic curves.

1.4.1 Elliptic curves

By definition, an *elliptic curve* over $\mathbb{C}$ is a curve described by an equation of the form $y^2 = f(x)$, where $f(x) \in \mathbb{C}[x]$ is a cubic polynomial with distinct roots. By making a suitable linear change of variables we can assume that the curve has the *Weierstrass form*

$$E = E_{a,b}: \qquad y^2 = 4x^3 - ax - b, \quad \text{where } \Delta_E := a^3 - 27b^2 \neq 0; \tag{1.57}$$

here we have used the fact that the polynomial $f(x) = 4x^3 - ax - b$ has distinct roots if and only if its discriminant $\mathrm{disc}(f) = 16(a^3 - 27b^2) \neq 0$.

Note that the change of variables $x_1 = \lambda^2 x$, $y_1 = \lambda^3 y$, where $\lambda \in \mathbb{C}^\times$, transforms the above elliptic curve to the (isomorphic) elliptic curve

$$E_1 = E_{a_1,b_1}: \qquad y_1^2 = 4x_1^3 - a_1 x - b_1,$$

in which $a_1 = a/\lambda^4$ and $b_1 = b/\lambda^6$; thus $\Delta_{E_1} = \Delta_E \cdot \lambda^{-12}$. In particular, the discriminant Δ_E is not preserved under isomorphisms of elliptic curves. However, it is immediate that

$$j_E \overset{\text{def}}{=} \frac{(12a)^3}{\Delta_E} = \frac{(12a)^3}{a^3 - 27b^2}$$

is invariant under such transformations; this number j_E is called the *j-invariant* of E.

It is often useful to "compactify" E by adding a point $P_\infty = (\infty, \infty)$ to E. In fact, $\overline{E} = E \cup \{P_\infty\}$ has a natural *group structure* (with identity P_∞) where the addition is given by the so-called *chord-tangent* method; cf. e.g. [ST], p. 15ff. for more details.

Remark. In many texts (such as [ST]) one finds in place of the (classical) Weierstrass form (as above) the equivalent form

$$\tilde{E}_{A,B}: \qquad Y^2 = X^3 + AX + B, \quad \text{where } 4A^3 + 27B^2 \neq 0,$$

which has certain advantages. Note that $\tilde{E}_{A,B}$ is obtained from $E_{a,b}$ by the transformation $x = X$ and $y = 2Y$, and so $A = -a/4$, $B = -b/4$ and $\Delta_{\tilde{E}_{A,B}} := -16(4A^3 + 27B^2) = a^3 - 27b^2 = \Delta_{E_{a,b}}$. Moreover, its j-invariant is $j_{\tilde{E}_{A,B}} := -\frac{(48A)^3}{\Delta_{\tilde{E}_{A,B}}} = j_{E_{a,b}}$.

The above is an *algebraic description* of elliptic curves (which in fact can be generalized to any field K of characteristic $\neq 2, 3$ in place of $\mathbb{C}$). However, for complex elliptic curves we also have an *analytic description:* there is a (unique) lattice $L \subset \mathbb{C}$ such that E "equals" $\mathbb{C}/L$. This identification is obtained by the theory of *doubly periodic (or elliptic) functions*, which we consider next.

1.4.2 Elliptic functions

Since the basic theory of such functions is presented in most standard texts on complex analysis (cf. e.g. Ahlfors[Ah], chapter 7), we shall only briefly recall the main facts.

Let $L \subset \mathbb{C}$ be a *lattice* in $\mathbb{C}$; thus $L = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$, where $\omega_1/\omega_2 \notin \mathbb{R}$. Note that by interchanging ω_1 and ω_2 if necessary, we can always assume that $\mathrm{Im}(\omega_1/\omega_2) > 0$, i.e. that $\omega_1/\omega_2 \in \mathfrak{H}$, and we shall do so tacitly in the sequel.

An *elliptic or doubly-periodic function* with period lattice L is a meromorphic function f defined on $\mathbb{C}$ such that

$$f(z+\omega) = f(z), \quad \text{for all } \omega \in L.$$

Since there are no non-constant *holomorphic* elliptic functions (cf. Ahlfors[Ah], p. 262 or [Ko], p. 15), we must allow poles. The simplest non-constant elliptic function is the Weierstrass $\wp$-function,

$$\wp_L(z) = \frac{1}{z^2} + \sum_{\omega \in L}{}' \left[\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right],$$

which has a pole of order 2 at every lattice point $\omega \in L$. By expanding $\frac{1}{(z-\omega)}$ as a power series and rearranging the terms, we obtain the following Laurent series expansion of $\wp_L$ (in a neighbourhood of 0):

$$\wp_L(z) = \frac{1}{z^2} + \sum_{k=2}^{\infty} (k+1) G_{k+2}(L) z^k, \quad \text{where } G_k(L) = \sum_{\omega \in L}{}' \frac{1}{\omega^k} \tag{1.58}$$

Note that this function $G_k(L)$ is closely related to the function $G_k(\tau)$ which was defined earlier in subsection 1.2.1; in fact, we have

$$G_k(L) = \sum_{m,n \in \mathbb{Z}}{}' \frac{1}{(m\omega_1 + n\omega_2)^k} = \frac{1}{\omega_2^k} \sum_{m,n \in \mathbb{Z}}{}' \frac{1}{(m(\omega_1/\omega_2) + n)^k} = \omega_2^{-k} G_k(\omega_1/\omega_2),$$

and so in particular, $G_k(\mathbb{Z}\tau + \mathbb{Z}) = G_k(\tau)$. Thus, we see that the coefficients of the expansion (1.58) of $\wp_L$ are given by modular forms!

By differentiating (1.58), one deduces easily that $\wp$ satisfies the differential equation

$$(\wp_L')^2 = 4\wp_L^3 - g_2(L)\wp_L - g_3(L), \tag{1.59}$$

where $g_2(L) = 60G_4(L)$ and $g_3(L) = 140G_6(L)$; cf. [Ah], p. 268 or [Ko], p. 23. Thus we see that the assignment $z \mapsto \phi_L(z) := (\wp_\tau(z), \wp_\tau'(z))$ defines a map from $\mathbb{C}/L$ to the plane cubic curve $E = E_L : y^2 = 4x^3 - g_2(L)x - g_3(L)$, which is in fact an elliptic curve because $\Delta_E = g_2(L)^3 - 27g_3(L)^2 = \Delta(L) := \omega_2^{-12}\Delta\left(\frac{\omega_1}{\omega_2}\right) \neq 0$ (since $\Delta(\tau)$ does not vanish on the upper half plane (cf. (1.39)). Note that the j-invariant of E_L is

$$j(L) := j_{E_L} = \frac{(12g_2(L))^3}{\Delta_{E_L}} = j\left(\frac{\omega_1}{\omega_2}\right), \tag{1.60}$$

where $j(\tau)$ denotes the j-function of subsection 1.2.3. Moreover, we have:

Proposition 1.19 *The map $z \mapsto (\wp(z), \wp'(z))$ induces a bijection (in fact, an analytic isomorphism)*

$$\phi_L : \mathbb{C}/L \xrightarrow{\sim} \overline{E}_L = \overline{E}_{g_2(L),g_3(L)}.$$

In addition, this is an isomorphism of groups.

Proof. The first assertion is easily verified; cf. [Ko], p. 24. The second assertion (about the group laws) is just a restatement of the *addition law* of the Weierstrass $\wp$-function (cf. [Ah], p. 269 or [Ko], p. 34.)

Remarks. 1) The *inverse* of the map ϕ_L is given by the (multi-valued) integral

$$\phi_L^{-1}(P) = \int_{P_\infty}^{P} \frac{dz}{\sqrt{f(z)}} + L \in \mathbb{C}/L,$$

where $f(z) = 4z^3 - g_2(L)z - g_3(L)$, and where the integral is over any path (on $\overline{E}$) which joins P_∞ to P; such an integral is (essentially) what is called an *elliptic integral*.

Historically, it was the study of elliptic integrals that gave birth to elliptic functions and to elliptic curves. In fact, the study of elliptic integrals begins with Fagnano's discovery (1718) that there is a simple formula for doubling the arc length $s(r) = \int_0^r \frac{dt}{\sqrt{1-t^4}}$ of a lemniscate, i.e. to find u such that $s(u) = 2s(r)$; cf. Siegel[Si], p. 1ff for the precise details. In 1753 Euler discovered that Fagnano's observation is a special case of a general *addition law* for the lemniscate integral $s(r)$, i.e. that there is a simple formula for the solution u of $s(u) = s(r) + s(r')$ in terms of r and r', and subsequently he (and Legendre) noticed that this is true more generally for all *elliptic integrals*

$$I_f(r) \;=\; \int_0^r \frac{dx}{\sqrt{f(x)}},$$

where $f(x)$ is any cubic or quartic polynomial. Later, Weierstrass discovered his $\wp$-function in his (successful) attempt to invert elliptic integrals. In particular, the addition law for the Weierstrass $\wp$-function is just a restatement of Euler's addition formula for elliptic integrals.

2) The Weierstrass function $\wp_L$ actually gives rise to *all* elliptic functions as follows. First of all, every *even* elliptic function with period lattice L is rational functions in $\wp_L$, i.e. the set $\mathcal{M}(L)^+$ of all even L-periodic elliptic functions is the field $\mathcal{M}(L)^+ = \mathbb{C}(\wp_L)$ of rational functions in $\wp_L$. Moreover, the set $\mathcal{M}(L)$ of all L-periodic elliptic functions is the field generated by $\wp_L$ and by its derivative $\wp'(z)$, i.e.

$$\mathcal{M}(L) = \mathbb{C}(\wp_L, \wp'_L),$$

i.e. every elliptic function is a rational function in $\wp$ and $\wp'$; cf. [Ko], p. 18. Note that the differential equation (1.59) shows that $\mathcal{M}(L) = \mathbb{C}(\wp_L, \wp'_L)$ is a quadratic extension of $\mathcal{M}(L)^+ = \mathbb{C}(\wp_L)$.

1.4.3 Lattice functions

In the previous subsection on elliptic functions we saw that certain modular forms miraculously appeared as values attached to lattices, i.e. as lattice functions; in particular, these modular forms extend to lattice functions. This is in fact no accident, for it turns out that there is a *complete dictionary* between lattice functions (of weight k) and functions on the upper half plane of weight k with respect to $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, as we shall see presently. For this, we first introduce the following definitions and notations.

Definition. Let $\mathcal{L} = \{L \subset \mathbb{C}\}$ denote the set of all lattices in $\mathbb{C}$. A *lattice function of weight k* is a function $F : \mathcal{L} \to \mathbb{C}$ such that

$$F(cL) \;=\; c^{-k}F(L), \quad \forall c \in \mathbb{C}^\times.$$

We denote the set of such functions by $\mathcal{F}_k(\mathcal{L})$. Moreover, a function $f : \mathfrak{H} \to \mathbb{C}$ is said to be *of weight k* with respect to Γ if it satisfies the transformation rule (1.2). The set of all such functions is denoted by $\mathcal{F}_k(\mathfrak{H}, \Gamma)$, i.e.

$$\mathcal{F}_k(\mathfrak{H}, \Gamma) = \{f : \mathfrak{H} \to \mathbb{C} \text{ with } f|_k\gamma = f, \text{ for all } \gamma \in \Gamma\}.$$

Example. The lattice function G_k defined in (1.58) has weight k (i.e. $G_k \in \mathcal{F}_k(\mathcal{L})$) because

$$G_k(cL) \;=\; \sum_{\omega \in cL}{}' \frac{1}{\omega^k} \;=\; \sum_{\omega \in L}{}' \frac{1}{(c\omega)^k} \;=\; \frac{1}{c^k}\sum_{\omega \in L}{}' \frac{1}{\omega^k} \;=\; c^{-k}G_k(L).$$

We now prove:

Proposition 1.20 *(a) The map $L : \mathfrak{H} \to \mathcal{L}$ given by $L(\tau) = \mathbb{Z}\tau + \mathbb{Z}$ induces a bijection*

$$\overline{L} : \Gamma\backslash\mathfrak{H} \;\xrightarrow{\sim}\; \mathcal{L}/\mathbb{C}^\times.$$

(b) The pull-back map $F \mapsto L^*F = F \circ L$ induces for each k a bijection

$$L^* : \mathcal{F}_k(\mathcal{L}) \;\xrightarrow{\sim}\; \mathcal{F}_k(\mathfrak{H}, \Gamma)$$

between the set $\mathcal{F}_k(\mathcal{L})$ of lattice functions of weight k and the set $\mathcal{F}_k(\mathfrak{H}, \Gamma)$ of functions on $\mathfrak{H}$ which have weight k with respect to Γ.

Proof. (a) It is immediate that the rule $\tau \mapsto \overline{L}(\tau) := L(\tau)\mathbb{C}^\times$ defines a surjection $\overline{L} : \mathfrak{H} \to \mathcal{L}/\mathbb{C}^\times$ because any lattice $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 \in \mathcal{L}$ can be written as $\Lambda = L(\omega_1/\omega_2)\omega_2$. Moreover, $\overline{L}$ is constant on Γ-orbits and hence defines a surjection $\overline{L} : \Gamma\backslash\mathfrak{H} \to \mathcal{L}/\mathbb{C}^\times$ because if $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$, then $L(g(\tau)) = (\mathbb{Z}(a\tau+b) + \mathbb{Z}(c\tau+d))(c\tau+d)^{-1} = L(\tau)(c\tau + d)^{-1}$. Here we have used one direction of the following easy general fact:

Fact. If $\underline{\omega} = (\omega_1, \omega_2), \underline{\omega}' = (\omega_1', \omega_2') \in \mathcal{B} := \{(\omega_1, \omega_2) \in \mathbb{C}^2 : \omega_1/\omega_2 \in \mathfrak{H}\}$, then

$$\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 \;=\; \mathbb{Z}\omega_1' + \mathbb{Z}\omega_2' \quad \Leftrightarrow \quad \exists g \in \Gamma = \mathrm{SL}_2(\mathbb{Z}) \,:\, g\underline{\omega}^t = (\underline{\omega}')^t. \tag{1.61}$$

[Indeed, by linear algebra we have that $\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 = \mathbb{Z}\omega_1' + \mathbb{Z}\omega_2' \Leftrightarrow \exists g \in \Gamma = \mathrm{GL}_2(\mathbb{Z}) : g\underline{\omega}^t = (\underline{\omega}')^t$. Moreover, if this is the case, then $g(\frac{\omega_1}{\omega_2}) = \frac{\omega_1'}{\omega_2'}$ (viewing g as a fractional linear transformation). But since $\frac{\omega_1}{\omega_2}, \frac{\omega_1'}{\omega_2'} \in \mathfrak{H}$, this forces that $\det(g) > 0$, i.e. that $g \in \mathrm{SL}_2(\mathbb{Z})$.]

It remains to show that $\overline{L} : \Gamma\backslash\mathfrak{H} \to \mathcal{L}/\mathbb{C}^\times$ is injective. Thus, suppose $\tau, \tau' \in \mathfrak{H}$ are such that $L(\tau)\lambda = L(\tau')$ for some $\lambda \in \mathbb{C}^\times$. Then by (1.61) $\exists g \in \mathrm{SL}_2(\mathbb{Z})$ such that $g(\lambda\tau, \lambda) = (\tau', 1)$. But then $g(\tau) = g(\frac{\lambda\tau}{\lambda}) = \frac{\tau'}{1} = \tau'$ (viewing g as a fractional linear transformation), and so $\tau' \in \Gamma\tau = \mathrm{orbit}_\Gamma(\tau)$. Thus $\overline{L}$ is injective and hence bijective.

(b) First note that if $f \in \mathcal{F}(\mathfrak{H})$ is any ($\mathbb{C}$-valued) function on $\mathfrak{H}$, then the rule

$$h_k(f)(\omega_1, \omega_2) := \omega_2^{-k} f(\omega_1/\omega_2)$$

defines a $\mathbb{C}$-valued map $h_k(f) \in \mathcal{F}(\mathcal{B})$ on the set $\mathcal{B} = \{(\omega_1, \omega_2) \in \mathbb{C}^2 : \omega_1/\omega_2 \in \mathfrak{H}\}$ of (oriented) bases of lattices. Now clearly $h_k(f)(\lambda\underline{\omega}) = \lambda^{-k} h_k(\underline{\omega}), \forall \lambda \in \mathbb{C}^\times$ and $\underline{\omega} = (\omega_1, \omega_2) \in \mathcal{B}$, i.e. $h_k(f) \in \mathcal{F}_k$ has weight k. Thus, the homogenization map h_k defines a map and bijection

$$h_k : \mathcal{F}(\mathfrak{H}) \ \tilde{\to}\ \mathcal{F}_k(\mathcal{B})$$

between the set of functions on $\mathfrak{H}$ and the set of functions on $\mathcal{B}$ of weight k.

We next observe that the linear action of $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{C}^2$ induces an action on $\mathcal{B} \subset \mathbb{C}^2$, and a short computation shows that

$$h_k(f) \circ g \ = \ h_k(f|_k g), \quad \text{for all } g \in \Gamma. \tag{1.62}$$

Thus we see that h_k defines a bijection

$$h_k : \mathcal{F}_k(\mathfrak{H}, \Gamma) \ \tilde{\to}\ \mathcal{F}_k(\mathcal{B})^\Gamma = \mathcal{F}_k(\Gamma\backslash\mathcal{B})$$

between the set $\mathcal{F}_k(\mathfrak{H}, \Gamma)$ of functions on $\mathfrak{H}$ of weight k with respect to Γ and the set $\mathcal{F}_k(\mathcal{B})^\Gamma$ of Γ-invariant functions of weight k on $\mathcal{B}$; note that the latter can be identified with the set $\mathcal{F}_k(\Gamma\backslash\mathcal{B})$ of functions of weight k on the quotient $\Gamma\backslash\mathcal{B}$.

On the other hand, we have by (1.61) a natural identification $\Gamma\backslash\mathcal{B} \tilde{\to} \mathcal{L}$ given by $(\omega_1, \omega_2) \mapsto \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ and so we see that h_k defines a bijection $h_k : \mathcal{F}_k(\mathfrak{H}, \Gamma) \tilde{\to} \mathcal{F}_k(\mathcal{L})$ which is given by the rule

$$h_k(f)(\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2) = \omega_2^{-k} f(\omega_1/\omega_2).$$

Clearly, the inverse of this bijection is the map $L^* : F \mapsto L^*F$, and so L^* is a bijection, as claimed.

Remark. The notion of a lattice function is (via the above Fact) closely related to the concept of a *homogeneous modular form* which is frequently found in the classical literature; cf. [Sch], p. 38. The above approach via lattices may be found in [Se1], p. 81.

1.4.4 The moduli space $\mathfrak{M}_1$

As we saw in subsection 1.4.2, the theory of elliptic functions shows that every lattice $L \subset \mathbb{C}$ gives rise to an elliptic curve $E_L \subset \mathbb{C} \times \mathbb{C}$ and that we have an identification $\mathbb{C}/L \simeq \overline{E}_L$. In fact, *every* elliptic curve $E \subset \mathbb{C} \times \mathbb{C}$ (in Weierstrass form) arises in this way, as we shall see presently. For this, we shall first prove:

Proposition 1.21 *The modular function $j \in \mathbf{A}_0$ induces an isomorphism*

$$j : \Gamma\backslash\mathfrak{H} \to \mathbb{C}.$$

Thus, for every $c \in \mathbb{C}$ there exists a lattice L, unique up to scaling, such that $j(L) = c$.

Proof. Since j is a modular function of weight 0 without a pole on $\mathfrak{H}$, it induces a map $j : \Gamma\backslash\mathfrak{H} \to \mathbb{C}$. To show that j is bijective, let $c \in \mathbb{C}$. Then $j - c \in \mathbf{A}_0$ and $v_\infty(j - c) = v_\infty(j) = -1$. Thus, by (1.38) we know that $j - c$ has a unique zero in $\Gamma\backslash\mathfrak{H}$, i.e. there is a unique point $\tau \in \Gamma\backslash\mathfrak{H}$ such that $j(\tau) = c$. Thus $j : \Gamma\backslash\mathfrak{H} \to \mathbb{C}$ is bijective. This proves the first assertion, and the second follows from Proposition 1.20(a) (together with formula (1.60)).

We can now refine the above result to show that each (Weierstrass) elliptic curve $E \subset \mathbb{C} \times \mathbb{C}$ comes from a unique lattice $L \subset \mathbb{C}$.

Proposition 1.22 *For any $a, b \in \mathbb{C}$ with $a^3 \neq 27b^2$ there is a unique lattice L such that $g_2(L) = a$ and $g_3(L) = b$. Thus, the map $L \mapsto E_L = E_{g_2(L), g_3(L)}$ defines a bijection*

$$\Phi : \mathcal{L} \xrightarrow{\sim} \{E_{a,b} \subset \mathbb{C} \times \mathbb{C}\}$$

between the set $\mathcal{L} = \{L \subset \mathbb{C}\}$ of lattices in $\mathbb{C}$ and the set of Weierstrass curves in $\mathbb{C} \times \mathbb{C}$.

Proof. Suppose first that $a = 0$. By (1.40) we know that $g_2(\rho) = 0$, and $g_3(\rho) \neq 0$, so if we choose $\lambda \in \mathbb{C}$ such that $\lambda^6 = g_3(\rho)b^{-1}$, then $L = \lambda(\mathbb{Z}\rho + \mathbb{Z})$ satisfies $g_2(L) = 0$ and $g_3(L) = \lambda^{-6} g_3(\mathbb{Z}\rho + \mathbb{Z}) = \lambda^{-6} g_3(\rho) = b$.

Next, assume that $a \neq 0$ and put $c = \frac{(12a)^3}{a^3 - 27b^2} \neq 0$. Then $b^2 = \frac{(c-12^3)}{27c} a^3$. By Proposition 1.21 there exists a lattice L_0 such that $j(L_0) = c$; thus $g_3(\lambda L_0)^2 = \frac{(c-12^3)}{27c} g_2(\lambda L_0)^3$, for any $\lambda \in \mathbb{C}^\times$. Choose λ such that $\lambda^4 = g_2(L_0)a^{-1}$, and put $L_1 = \lambda L_0$. Then $g_2(L_1) = \lambda^{-4} g_2(L_0) = a$. Moreover, $g_3(L_1)^2 = \frac{(c-12^3)}{27c} g_2(L_1)^3 = \frac{(c-12^3)}{27c} a^3 = b^2$, so $g_3(L_1) = \pm b$. If $g_3(L_1) = b$, then take $L = L_1$; otherwise, take $L = iL_1$.

It remains to show that L is uniquely determined by a and b. If $a = 0$, then $j(L) = 0$, and so $L = \lambda(\mathbb{Z}\rho + \mathbb{Z})$, for some $\lambda \in \mathbb{C}^\times$. Then $g_3(L) = b \Leftrightarrow \lambda^6 = g_3(\rho)b^{-1}$, and so λ is unique up to a sixth root of unity, i.e. up to a power of $(-\rho)$. But $(-\rho)^k(\mathbb{Z}\rho + \mathbb{Z}) = \mathbb{Z}\rho + \mathbb{Z}$, so L is uniquely determined by this property.

Next, suppose $b = 0$. Then an analogous argument shows that $L = \lambda(\mathbb{Z}i + \mathbb{Z})$, where $\lambda^4 = g_2(i)a^{-1}$, so λ is unique up to a power of i, and hence L is unique.

Finally, suppose that $ab \neq 0$, and that L_1, L_2 are two lattices such that $g_2(L_1) = g_2(L_2) = a$ and $g_3(L_1) = g_3(L_2) = b$. Then also $j(L_1) = j(L_2)$, and so $L_2 = \lambda L_1$, for some $\lambda \in \mathbb{C}^\times$. Thus $a = g_2(L_2) = \lambda^{-4} g_2(L_1) = \lambda^{-4} a$, and so $\lambda^4 = 1$, and similarly $\lambda^6 = 1$, and hence $\lambda^2 = \lambda^6/\lambda^4 = 1$. Thus $\lambda = \pm 1$, and so $L_1 = L_2$.

We thus see that every elliptic curve E is "uniformized" by a unique lattice L. This fact can be used to prove the following purely algebraic statement about isomorphism classes of elliptic curves:

Corollary 1.23 *Two elliptic curves E_1 and E_2 are isomorphic if and only if they have the same j-invariant, i.e.*

$$E_1 \simeq E_2 \quad \Leftrightarrow \quad j_{E_1} = j_{E_2}.$$

Proof. If $E_1 \simeq E_2$, then clearly $j_{E_1} = j_{E_2}$, as was mentioned earlier in subsection 1.4.1.

Conversely, suppose that $j_{E_1} = j_{E_2}$, and write $E_k = E_{a_k, b_k}$, for $k = 1, 2$. Then by Proposition 1.22 there exist lattices L_k such that $g_2(L_k) = a_k$ and $g_3(L_k) = b_k$, for $k = 1, 2$. Now by hypothesis and formula (1.60) we have $j(L_1) = j_{E_1} = j_{E_2} = j(E_2)$, and so by Proposition 1.21 it follows that $L_2 = \lambda L_1$, for some $\lambda \in \mathbb{C}^\times$. Then $a_2 = g_2(\lambda L_1) = \lambda^{-4} g_2(L_1) = \lambda_{-4} a_1$, and similarly $b_2 = \lambda^{-6} b_1$. Thus, the map $(x, y) \mapsto (\lambda^2 x, \lambda^3 y)$ defines an isomorphism $E_1 \xrightarrow{\sim} E_2$.

We thus see that the j-invariant completely characterizes an elliptic curve up to isomorphism, i.e. j establishes a bijection between the set $\mathbb{C}$ and the set

$$\mathfrak{M}_1 = \{E_{a,b} : a, b \in \mathbb{C}, a^3 \neq 27b^2\}/\simeq$$

of *isomorphism classes* of elliptic curves. More precisely, we have:

Theorem 1.24 *The modular function j factors over $\mathfrak{M}_1$ to induce isomorphisms*

$$\Gamma\backslash\mathfrak{H} \xrightarrow{\sim} \mathcal{L}/\mathbb{C}^\times \xrightarrow{\sim} \mathfrak{M}_1 \xrightarrow{\sim} \mathbb{C}$$

which are induced by the maps $\tau \mapsto L(\tau) = \mathbb{Z}\tau + \mathbb{Z}$, $L \mapsto E_L$, and $E \mapsto j_E$, respectively.

Proof. The fact that the composition of these maps is j is the content of formula (1.60). Now $j : \Gamma\backslash\mathfrak{H} \to \mathbb{C}$ is an isomorphism by Proposition 1.21, and the first and third maps are isomorphisms by Proposition 1.20(a) and Corollary 1.23, respectively. Thus, all maps are isomorphisms.

Remark. The above set $\mathfrak{M}_1$ is called *moduli space of elliptic curves*. Thus, the above results show that this "abstract" set has a natural algebraic/analytic structure because we can identify it (via j) with the complex plane $\mathbb{C}$. (Note that $\mathbb{C}$ is also an algebraic object because it can be identified with the set of points of the affine line $\mathbb{A}^1_{\mathbb{C}}$.)

More generally, a *moduli problem* attempts to classify isomorphism classes of algebraic objects by identifying them with the points of an algebraic/analytic object. We will encounter further moduli problems when we discuss modular forms of higher level (cf. chapter **??**).

1.5 Hecke Operators

In section 1.2 we encountered the (normalized) discriminant function

$$\Delta_1(z) = \frac{1}{1728}(E_4^3 - E_6^2) = \sum_{n\geq 1} \tau(n) q^n,$$

where $\tau(n)$ is the Ramanujan τ-function. As was mentioned, Ramanujan conjectured in 1916 that

1) $\tau(n)$ is multiplicative;

2) $\tau(p^r)$ can be expressed in terms of $\tau(p^{r-1})$, $\tau(p^{r-2})$ and $\tau(p)$ for $r \geq 2$.

This conjecture was then subsequently proved by Mordell in (1917). Later in 1937, Hecke addressed the following questions in his fundamental paper[He]:

Questions: 1) *Are there any other modular forms whose Fourier coefficients are multiplicative? How many such modular forms are there?*

2) *Is there an analogue of property* 2) *above?*

In his paper, Hecke made the following remarkable discoveries:

1) There are at most $\dim \mathbf{M}_k$ (non-zero) modular forms f of weight k whose Fourier coefficients are multiplicative.

2) If f is as in 1), then its Fourier coefficients $a_{p^r}(f)$ for prime powers satisfy a recursion relation which is similar to that of the $\tau(p^r)$'s; cf. Corollary 1.37 below.

Two year later, Hecke's student H. Petersson[Pe] was able to extend and complete Hecke's work by proving (cf. Corollary 1.40 below):

1)* There are precisely $\dim \mathbf{M}_k$ (non-zero) modular forms f of weight k whose Fourier coefficients are multiplicative, and these form a basis of $\mathbf{M}_k$.

Hecke's idea was to make use of certain operators T_n (now called *Hecke Operators*) which act on modular forms. Although these and other operators had been studied earlier by Kronecker, Klein, Gierster and Hurwitz in their study of *modular correspondences*, it was Hecke who first realized their importance in connection with modular forms. In addition, he discovered that these operators satisfy some basic relations which then force relations on the Fourier coefficients of forms.

These Hecke Operators will be defined and studied in the next subsection. Then in subsection 1.5.2 we shall prove the *key result* of Hecke that a modular form $f \in \mathbf{M}_k$ is a simulataneous eigenfunction of all the Hecke operators if and only if its (normalized) Fourier coefficients are multiplicative. Finally, in subsection 1.5.3 we shall show, using the so-called *Petersson inner product*, that such eigenfunctions form a basis of $\mathbf{M}_k$.

1.5.1 The Hecke Algebra

Let $f \in \mathbf{M}_k$ be a modular form of weight k. For each integer $n \geq 1$, put

$$(f|_k T_n)(z) = (T_n f)(z) \;=\; n^{k-1} \sum_{\substack{a \geq 1 \\ ad = n \\ 0 \leq b < d}} d^k f\left(\frac{az+b}{d}\right) \tag{1.63}$$

This can also be written more intrinsically in terms of lattice functions as follows. Given a lattice function $F \in \mathcal{F}_k(\mathcal{L})$ of weight k (cf. section 1.4.3), define for each integer $n \geq 1$ the map $\mathcal{T}_n : \mathcal{F}_k(\mathcal{L}) \to \mathcal{F}_k(\mathcal{L})$ be the formula

$$\mathcal{T}_n(F)(L) \;=\; n^{k-1} \sum_{\substack{L' \subset L \\ [L : L'] = n}} F(L'), \tag{1.64}$$

where the sum extends over all sublattices L' of L of index n. Then one easily shows (cf. Serre[Se1], p. 100) that we have:

$$T_n(L^* F) = L^*(\mathcal{T}_n(F)), \tag{1.65}$$

where $L^* : \mathcal{F}_k(\mathcal{L}) \xrightarrow{\sim} \mathcal{F}_k(\mathfrak{H}, \Gamma)$ is the map which identifies the set $\mathcal{F}_k(\mathcal{L})$ of lattice functions of weight k with the set $\mathcal{F}_k(\mathfrak{H}, \Gamma)$ of functions on $\mathfrak{H}$ of weight k with respect to Γ; cf. Proposition 1.20.

Proposition 1.25 *Each Hecke operator T_n maps modular forms to modular forms and cusp forms to cusp forms of the same weight. Moreover, if $f = \sum a_n(f) q^n \in \mathbf{M}_k$, then the Fourier coefficients of $f|_k T_n$ are given by*

$$a_m(f|_k T_n) = \sum_{d|(m,n)} d^{k-1} a_{mn/d^2}(f), \quad \textit{for all } m \geq 0; \tag{1.66}$$

in particular,

$$a_0(f|_k T_n) = \sigma_{k-1}(n) a_0(f), \quad a_1(f|_k T_n) = a_n(f),$$

and, if $n = p$ is prime,

$$a_m(f|_k T_p) = \begin{cases} a_{mp}(f) & \textit{if } p \nmid m \\ a_{mp}(f) + p^{k-1} a_{m/p}(f) & \textit{if } p \mid m \end{cases}$$

Proof. (Sketch) Let $f \in \mathbf{M}_k$. Then by (1.63) we see that $f|_k T_n$ is holomorphic on $\mathfrak{H}$, and by (1.64) (and (1.65)) we see (using Proposition 1.20) that $f|_k T_n$ is weakly modular of weight k for Γ. Finally, a short computation (cf. [Se1], p. 100) shows that $f|_k T_n$ has Fourier expansion $f|_k T_n(z) = \sum_{m \geq 0} a_{m,n} q^n$ where the $a_{m,n}$'s are given by (1.66). Thus $f|_k T_n$ is holomorphic at infinity, and so $f|_k T_n \in \mathbf{M}_k$. Note that (1.66) also shows that if $f \in \mathbf{S}_k$ (i.e. $a_0(f) = 0$), then $f|_k T_n \in \mathbf{S}_k$.

The Hecke operators T_n satisfy the following fundamental relations which therefore induce relations on the Fourier coefficients of modular forms, as we shall see below.

Theorem 1.26 *As linear operators on* $\mathbf{M}_k$*, the Hecke operators satisfy the relations*

$$T_m T_n = T_{mn} \quad \textit{for all integers } m, n \geq 1 \textit{ with } (m,n) = 1. \tag{1.67}$$

$$T_p T_p^r = T_{p^{r+1}} + p^{k-1} T_{p^{r-1}}, \quad \textit{if } p \textit{ is a prime and } r \geq 1. \tag{1.68}$$

Proof. (Sketch) By (1.65), it is enough to verify the corresponding properties for the $\mathcal{T}_n$'s, and these follow easily from the definition (1.64) and properties of lattices. For example, if $(m,n) = 1$, then each sublattice L' of a lattice L of index $[L : L'] = mn$ is uniquely the intersection of two intermediate lattices L'_1,L'_2 of index $[L : L'_1] = m$ and $[L : L'_2] = n$, from which (1.67) follows readily. See [Se1], Chapter VII, Proposition 10 and 11 (p. 98) for more details.

Definition. The *Hecke algebra* $\mathbb{T} = \mathbb{T}_k \subset \mathrm{End}_{\mathbb{C}}(\mathbf{M}_k)$ is the $\mathbb{C}$-algebra generated by all the operators T_n. Thus, by Theorem (1.26) we see that $\mathbb{T}$ is a commutative algebra which coincides with the $\mathbb{C}$-algebra generated by $T_1 = id$ and all the T_p's, where p is prime.

A modular form $f \in \mathbf{M}_k$ is called a $\mathbb{T}$-*eigenfunction with eigenvalues* $\{\lambda_n\}_{n\geq 1}$ if f satisfies the relations

$$f|_k T_n = \lambda_n f, \quad \forall n \geq 1. \tag{1.69}$$

If this the case, then there exists a unique $\mathbb{C}$-linear ring homomorphism $\chi_f : \mathbb{T} \to \mathbb{C}$ with $\chi_f(T_n) = \lambda_n$, for all $n \geq 1$, such that we have

$$f|_k T = \chi_f(T) f, \quad \forall T \in \mathbb{T}. \tag{1.70}$$

This map χ_f is called the *character* of $\mathbb{T}$ associated to the $\mathbb{T}$-eigenfunction f.

Example 1.27 (a) If $\dim_{\mathbb{C}} \mathbf{M}_k = 1$, then each modular form $f \in \mathbf{M}_k$ of weight k is a $\mathbb{T}$-eigenfunction for some set of eigenvalues $\{\lambda_n\}$. Similarly, if $\dim_{\mathbb{C}} \mathbf{S}_k = 1$, then each cusp form $f \in \mathbf{S}_k$ of weight k is a $\mathbb{T}$-eigenfunction. In particular, by Theorem 1.1 (and the Remark following it) we see that the following are $\mathbb{T}$-eigenfunctions:

$$E_4,\ E_6,\ E_8,\ E_{10},\ \Delta,\ E_{14},\ \Delta E_4,\ \Delta E_6,\ \Delta E_8,\ \Delta E_{10},\ \Delta E_{14}.$$

[Indeed, since $f|_k T_n \in \mathbf{M}_k$ by Proposition 1.25, we see that $f|_k T_n = \lambda_n f$, for some $\lambda \in \mathbb{C}$ (because $\dim \mathbf{M}_k = 1$), and so f is a $\mathbb{T}$-eigenfunction.]

(b) Each Eisenstein series E_k is a $\mathbb{T}$-eigenfunction with eigenvalues $\{\sigma_{k-1}(n)\}_{n\geq 1}$; cf. Example 1.34(a) below or Serre[Se1], p. 104.

The Fourier coefficients of a $\mathbb{T}$-eigenfunction f are closely related to its eigenvalues, as we shall now see. This implies that relations among the operators T_n induce relations among the Fourier coefficients of f.

Theorem 1.28 *If* $f = \sum a_n(f) q^n \in \mathbf{M}_k$ *is a* $\mathbb{T}$*-eigenfunction with eigenvalues* $\{\lambda_n\}_{n\geq 1}$*, then its Fourier coefficients satisfy the relations*

$$a_n(f) = a_1(f)\lambda_n, \quad \textit{for all } n \geq 1. \tag{1.71}$$

In particular, we have $a_1(f) \neq 0$ unless $f = 0$ (or $k = 0$). Moreover, if $f \neq 0$, then the Fourier coefficients $a_n = a_n(\tilde{f})$ of $\tilde{f} = f/a_1(f)$ satisfy:

$$a_{mn} = a_m a_n, \quad \textit{for all } (m,n) = 1, \tag{1.72}$$

$$a_{p^{r+1}} = a_p a_{p^r} - p^{k-1} a_{p^{r-1}}, \quad \textit{if } p \textit{ is a prime and } r \geq 1. \tag{1.73}$$

Proof. By (1.66) and (1.69) we have

$$a_n(f) \overset{(1.66)}{=} a_1(f|_k T_n) \overset{(1.69)}{=} a_1(\lambda_n f) = \lambda_n a_1(f),$$

which proves (1.71). From this we see that if $a_1(f) = 0$, then all $a_n(f)$'s are zero for $n \geq 1$ and so $f(z) = a_0(f)$ is constant. Thus either $f = 0$ or $k = 0$.

Now suppose $f \neq 0$, so $a_1(f) \neq 0$. Then for $(m,n) = 1$ we have by (1.67) that

$$\lambda_{mn}\tilde{f} \overset{(1.69)}{=} \tilde{f}|_k T_{mn} \overset{(1.67)}{=} (\tilde{f}|_k T_m)|_k T_n = (\lambda_m \tilde{f})|_k T_n = \lambda_n \lambda_m \tilde{f}.$$

Thus, $\lambda_{mn} = \lambda_m \lambda_n$. Since $a_n = a_n(\tilde{f}) = \lambda_n$, for all $n \geq 1$, we see that the a_n's are multiplicative, i.e. equation (1.72) holds.

The proof of equation (1.73) is analogous, using relation (1.68) in place of relation (1.67).

Remark. Let $f \in \mathbf{M}_k$ be a $\mathbb{T}$-eigenfunction which is *normalized* in the sense that $a_1(f) = 1$. Then $f = \tilde{f}$, and so (1.72) shows that its Fourier coefficients are multiplicative. Moreover, by (1.71) we see that they are the eigenvalues of $\mathbb{T}$, i.e. that we have $\lambda_n = a_n(f)$, for all $n \geq 1$.

Example 1.29 By Example 1.27 we know that the discriminant function Δ is a $\mathbb{T}$-eigenfunction of weight 12. Since $\Delta(z) = (2\pi)^{12} \sum \tau(n) q^n$ with $\tau(1) = 1$, it follows from (1.71) that its eigenvalues are $\{\tau(n)\}_{n \geq 1}$. Thus, by Theorem 1.28 we see that we have

$$\begin{aligned} \tau(mn) &= \tau(m)\tau(n), \quad \text{for all } (m,n) = 1, \\ \tau(p^{r+1}) &= \tau(p)\tau(p^r) - p^{11}\tau(p^{r-1}), \quad \text{if } r \geq 1 \text{ and } p \text{ is prime.} \end{aligned}$$

This, therefore, proves the first two of the three conjectures of Ramanujan mentioned in subsection 1.2.2.

Corollary 1.30 (Multiplicity 1) *If $f, g \in \mathbf{M}_k$, $k > 0$, are two non-zero $\mathbb{T}$-eigenfunctions with the same eigenvalues $\{\lambda_n\}_{n \geq 1}$, then $g = cf$, for some $c \in \mathbb{C}$.*

Proof. By Theorem 1.28 we know that $a_1(f) \neq 0$. Thus, putting $c = a_1(g)/a_1(f)$, we see that $a_n(g) = a_1(g)\lambda_n = c a_1(f)\lambda_n = c a_n(f)$, for all $n \geq 1$. Thus $g - cf = a_0(g) - c a_0(g)$ is a constant modular form of weight k, and hence is 0. This means that $g = cf$, as claimed.

1.5.2 L-functions

The relations satisfied by the Fourier coefficients of a $\mathbb{T}$-eigenfunction f are best understood in terms of the *Dirichlet series* or *L-function* associated to f.

Definition. If $f \in \mathbf{M}_k$, then its *associated L-function* is the Dirichlet series

$$L(f,s) = \sum_{n\geq 1} a_n(f) n^{-s}, \quad \text{where } f = \sum_{n\geq 0} a_n(f) q^n.$$

Observe that the constant term $a_0(f)$ of f is ignored in the definition of $L(f,s)$. Since by Corollary 1.14 we have $|a_n(f)| \leq cn^{k-1}$, for some $c > 0$, we see that $|L(f,s)| \leq c\sum_{n\geq 1} n^{k-1-s} = \zeta(s-k+1)$, and so the sum defining $L(f,s)$ converges absolutely for $Re(s) > k$.

Remark. The L-function $L(f,s)$ is closely related to its *Mellin-transform* $M(f,s)$ which is defined by

$$M(f,s) = \int_0^\infty (f(iy) - f(\infty)) y^s \frac{dy}{y}.$$

The precise relation is given by *Mellin's formula*

$$M(f,s) = (2\pi)^{-s}\Gamma(s)L(f,s), \tag{1.74}$$

which is easily derived (cf. Lang[La], p. 20). From this one concludes easily that $L(f,s)$ has an analytic continuation to $\mathbb{C}$ with at most a simple pole at $s = 1$. (In fact, $L(f,s)$ is holomorphic everywhere if $f \in S_k$ is a cusp form.)

Moreover, since f satisfies the transformation law $f(-1/z) = z^k f(z)$, its Mellin transform satisfies the functional equation

$$M(f,s) = (-1)^{k/2} M(f, k-s),$$

which therefore also gives the functional equation of $L(f,s)$.

Now Hecke showed in 1935/36 that the converse also holds: every L-function which has a functional equation and satisfies certain growth conditions comes from a modular form. More precisely:

Theorem 1.31 (Hecke) *Suppose f is a holomorphic function on $\mathfrak{H}$ which has a Fourier expansion $f(z) = \sum_{n=0} a_n q^n$ that converges absolutely and uniformly on each compact subset of $\mathfrak{H}$. Then $f \in \mathbf{M}_k$ if and only if the following conditions hold:*

(i) There is a $\nu > 0$ such that for $Im(z) \to 0$ we have $f(z) = O(Im(z)^{-s})$ (uniformly in $Re(s)$).

(ii) The Mellin transform $M(f,s) = (2\pi)^{-s}\Gamma(s)L(f,s)$ has an analytic continuation to the whole complex plane and satisfies the functional equation

$$M(f,s) = M(f, k-s) \tag{1.75}$$

and the function

$$M(f,s) + \frac{a_0}{s} + \frac{a_0}{k-s}$$

is holomorphic on the whole complex plane and is bounded in any vertical strip.

Proof. This is a special case of Theorem 7.2 of Iwaniec[Iw], p. 122, or of Theorem 4.3.5 of Miyake[Mi], p. 119.

Thus, the Dirichlet series $L(f,s)$ associated to a modular form $f \in \mathbf{M}_k$ always has an analytic continuation and a functional equation. However, in general it will not have an Euler product unless f is a $\mathbb{T}$-eigenfunction, as we shall see below in Theorem 1.35. A first step towards this is given by the following result:

Proposition 1.32 *Let $f \in \mathbf{M}_k$ be a modular form of weight k. If f is a $\mathbb{T}$-eigenfunction with eigenvalues $\{\lambda_n\}_{n\geq 1}$, i.e. if f satisfies (1.69), then*

$$L(f,s) = a_1(f) \prod_p \frac{1}{1 - \lambda_p p^{-s} + p^{k-1-2s}}. \tag{1.76}$$

Conversely, if $L(f,s)$ has an Euler product as above, then f is a $\mathbb{T}$-eigenfunction with eigenvalues $\{\lambda_n\}_{n\geq 1}$ given by (1.71). Thus

$$L(f,s) = a_1(f) \sum_{n\geq 1} \lambda_n n^{-s}.$$

Proof. This follows from Theorem 1.28 by using the following elementary lemma about Dirichlet series.

Lemma 1.33 *Let $\{a_n\}_{n\geq 1}$ be a sequence of complex numbers with $a_n = O(n^{k-1})$, for some k. Then*

$$\sum_{n=1}^{\infty} \frac{a_n}{n^s} \;=\; \prod_p \frac{1}{1 - a_p p^{-s} + p^{k-1-2s}}, \quad \textit{for } Re(s) > k.$$

if and only if the a_n's are multiplicative and satisfy the condition

$$a_p a_{p^r} = a_{p^{r+1}} + p^{k-1} a_{p^{r-1}}, \quad \textit{for every prime } p \textit{ and integer } r \geq 1.$$

Proof. Exercise.

Example 1.34 (a) The L-function associated to the Eisenstein series E_k for $k \geq 4$ is given by

$$L(E_k,s) = c_k \sum_{n\geq 1} \frac{\sigma_{k-1}(n)}{n^s} = c_k \zeta(s)\zeta(s-k+1).$$

Here the first equality is just the definition of the L-function (using (1.8)), and the second is a well-known identity of Dirichlet series. Indeed, any $a > 0$ we have

$$\zeta(s)\zeta(s-a) = \sum_{n=1}^{\infty} \frac{1}{n^s} \sum_{n=1}^{\infty} \frac{n^a}{n^s} = \sum_{n=1}^{\infty} \frac{1}{n^s} \sum_{d|n} d^a = \sum_{n=1}^{\infty} \frac{\sigma_a(n)}{n^s}.$$

Thus, since $\zeta(s)$ has an Euler product, so does $L(E_k, s)$; more precisely:

$$\begin{aligned} L(E_k, s) &= c_k \zeta(s)\zeta(s-k+1) = c_k \prod_p \frac{1}{(1-p^{-s})(1-p^{k-1-s})} \\ &= c_k \prod_p \frac{1}{1-\sigma_{k-1}(p)p^{-s}+p^{k-1-2s}}. \end{aligned}$$

Thus, by Proposition 1.32 we see that E_k is a $\mathbb{T}$-eigenfunction with eigenvalues $\{\sigma_{k-1}(n)\}_{n\geq 1}$.

(b) The discriminant function Δ is a $\mathbb{T}$-eigenfunction with eigenvalues $\{\tau(n)\}_{n\geq 1}$; cf. Example 1.29. The associated L-function is

$$L(\Delta, s) = (2\pi)^{12} \prod_p \frac{1}{1-\tau(p)p^{-s}+p^{11-2s}}.$$

The above theorem shows that if $f \in \mathbf{M}_k$ is a (normalized) $\mathbb{T}$-eigenfunction, then its Fourier coefficients $\{a_n(f)\}_{n\geq 1}$ are multiplicative. As Hecke showed, this property characterizes $\mathbb{T}$-eigenfunctions:

Theorem 1.35 (Hecke) *Let $f \in \mathbf{M}_k$ be a non-zero modular form of weight $k > 0$. Then its Fourier coefficients $\{a_n(f)\}_{n\geq 1}$ are multiplicative if and only if f is a normalized $\mathbb{T}$-eigenfunction with eigenvalues $\{a_n(f)\}_{n\geq 1}$.*

The proof of this theorem depends on the following lemma which is also of independent interest.

Lemma 1.36 *Let $f \in \mathbf{M}_k$, where $k > 0$ and let p be a prime. If $a_m(f) = 0$, for all $m \geq 1$ with $p \nmid m$, then $f = 0$.*

Proof. Put $f_0(z) = f(z/p)$. We now prove:

Claim 1. We have $f_0|_k g = f_0$, for all $g \in \Gamma^0(p) := \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma(1) : p|b\}$.

First note that $f_0 = p^k f|_k \alpha_p$, where $\alpha_p = \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$. Now if $g = \left(\begin{smallmatrix} a & bp \\ c & d \end{smallmatrix}\right) \in \Gamma^0(p)$, then $g = \alpha_p^{-1} g_1 \alpha_p$, with $g_1 = \left(\begin{smallmatrix} a & b \\ cp & d \end{smallmatrix}\right) \in \Gamma(1)$.Thus

$$p^{-k/2} f_0|_k g = f|_k \alpha_p g = f|_k g_1 \alpha_p = f|_k \alpha_p = p^{-k/2} f_0,$$

which proves claim 1.

Claim 2. $f_0 \in \mathbf{M}_k$.
Clearly f_0 is holomorphic on $\mathfrak{H}$. Moreover, since

$$f_0(z) = \sum_{n \geq 0} a_n(f) e^{2\pi i n z/p} \overset{\text{hypothesis}}{=} \sum_{\substack{n \geq 0 \\ p|n}} a_n(f) e^{2\pi i n z/p} = \sum_{n \geq 0} a_{np}(f) e^{2\pi i n z},$$

we see that f_0 is holomorphic at ∞ and that $f_0|T = f_0$, where $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Thus, by claim 1, $f_0|_k g = f_0, \forall g \in \langle T, \Gamma^0(p) \rangle$. But $\langle T, \Gamma^0(p) \rangle = \mathrm{SL}_2(\mathbb{Z})$ (exercise), so $f_0 \in \mathbf{M}_k$.
Claim 3. Put $f_j(z) := f_0(p^j z)$. Then $f_j \in \mathbf{M}_k$, for all $j \geq 0$.
We prove this by induction on j. Since $f_0 \in \mathbf{M}_k$ by claim 2 and $f_1 = f \in \mathbf{M}_k$ by hypothesis, we may assume $j \geq 2$. Now we note that

$$f_j|_k T_p = f_{j-1} + p^{k-1} f_{j+1}, \quad \text{if } j \geq 1,$$

because if $p \nmid m$, then $a_m(f) = 0 = a_m(f_{j-1} + p^{k-1} f_{j+1})$ whereas if $p|m$ then $a_m(f_j|_k T_p) = a_{mp}(f_j) + p^{k-1} a_{m/p}(f_j) = a_m(f_{j-1} + p^{k-1} f_{j+1})$ by (1.66). Thus, since $f_j, f_{j-1} \in \mathbf{M}_k$ by the induction hypothesis, we see that also $f_{j+1} = (f_j|_k T_p - f_{j-1})/p^{k-1} \in \mathbf{M}_k$.
Claim 4. If $f \neq 0$, then $f_0, f_1, f_2, \ldots$, are linearly independent.
If $f \neq 0$, then $m_j = \min\{m > 0 : a_m(f_j) \neq 0\} < \infty$. Then $m_j = p^j m_0$, for all $j \geq 0$, and so we see easily that the f_j's are linearly independent.

Now since $\dim_{\mathbf{M}_k} < \infty$, we see that claims 3 and 4 yield that $f = 0$, as desired.

Proof of Theorem 1.35. Since the Fourier coefficients of every normalized $\mathbb{T}$-eigenfunction are multiplicative by Theorem 1.28 (and the remark following it), it is enough to prove the converse.

Thus, suppose the Fourier coefficients of f are multiplicative. For a prime p, consider the function $f_p = f|_k T_p - a_p(f) f \in \mathbf{M}_k$. Then for every $m \geq 1$ with $p \nmid m$ we have $a_m(f_p) = a_{mp}(f) - a_p(f) a_m(f) = 0$ because the a_m's are multiplicative. Thus, f_p satisfies the hypotheses of Lemma 1.36 and so $f_p = 0$. This means $f|_k T_p = a_p(f) f$, and so f is a $\mathbb{T}$-eigenfunction. Moreover, by Theorem 1.28 we know that $a_1(f) \neq 0$. Thus, $a_1(f) = 1$ because by multiplicativity we have $a_1(f) = a_1(f) a_1(f)$. Thus f is normalized.

Corollary 1.37 *Let $f \in \mathbf{M}_k$ be a non-zero modular form whose Fourier coefficients $a_n = a_n(f)$ are multiplicative. Then its associated L-function has an Euler product of the form (1.76), and hence its Fourier coefficients satisfy the relation*

$$a_{p^{r+1}} = a_p a_{p^r} - p^{k-1} a_{p^{r-1}}, \quad \textit{for every prime } p \textit{ and integer } r \geq 1.$$

Proof. By Theorem 1.35 we know that f is a normalized $\mathbb{T}$-eigenfunction, and so the assertion follows from Theorem 1.28 and Lemma 1.33.

1.5.3 The Petersson Scalar Product

We now turn to study the *existence* of $\mathbb{T}$-eigenfunctions. Although we had already constructed some explicit examples, the previous theory does not allow us to prove the existence of sufficiently many $\mathbb{T}$-eigenfunctions, and so a new idea (due to H. Petersson) is necessary. This is based on the following concept.

Notation. For any two cusp forms $f, g \in \mathbf{S}_k$, put

$$\langle f, g \rangle = \int_D f(z)\overline{g(z)}y^{k-2}dxdy, \quad \text{where } z = x + iy.$$

It is easy to see that this integral converges and that it thus defines a (non-degenerate) hermitian pairing on $\mathbf{S}_k$, called the *Petersson scalar product*.

Proposition 1.38 *Each Hecke operator T_n is self-adjoint (or hermitian) with respect to the Petersson scalar product, i.e. we have $T_n^* = T_n$, or equivalently,*

$$\langle f|_k T_n, g\rangle = \langle f, g|_k T_n\rangle, \quad \textit{for all } f, g \in \mathbf{S}_k.$$

Proof. See [Ko], Proposition 48 (p. 171) or Miyake[Mi], Theorem 4.5.4 (p. 136).

We can use the Petersson product to prove the following fundamental result due to Petersson[Pe]:

Theorem 1.39 (Petersson) *The normalized $\mathbb{T}$-eigenfunctions of $\mathbf{M}_k$ form a basis of $\mathbf{M}_k$.*

Before proving this, let us review some basic linear algebra facts concerning (simultaneous) eigenvectors and eigenvalues of linear operators.

Review of Linear Algebra: Let V be a finite-dimensional $\mathbb{C}$-vector space and $\mathbb{T} \subset \mathrm{End}_{\mathbb{C}}(V)$ a *commutative* algebra of linear operators.

Recall: A non-zero vector $v \in V$ is called a (simultaneous) $\mathbb{T}$-*eigenvector* if for each $T \in \mathbb{T}$ there is a number $\chi(T) \in \mathbb{C}$ such that

$$T(v) = \chi(T)v.$$

Clearly, the map $T \mapsto \chi(T)$ defines a $\mathbb{C}$-linear ring homomorphism $\chi : \mathbb{T} \to \mathbb{C}$, i.e. a *character* of $\mathbb{T}$. (In particular, $\chi(id_V) = 1$.) Conversely, given a character χ, there exists at least one non-zero eigenvector $v \in V(\chi)$, i.e. the associated χ-*eigenspace*

$$V(\chi) = \{v \in V : T(v) = \chi(T)v, \text{ for all } T \in \mathbb{T}\}.$$

is non-zero: $V(\chi) \neq 0$. (This follows easily from the existence theorem of eigenvectors, using the fact that $\mathbb{T}$ is commutative.)

Facts: 1) If $\chi_1, \ldots, \chi_r$ are distinct characters of $\mathbb{T}$, then the associated eigenspaces are linearly independent, i.e.

$$\sum_{i=1}^{r} V(\chi_i) = \bigoplus_{i=1}^{r} V(\chi_i).$$

Thus, the set $\hat{\mathbb{T}} = \{\chi\}$ of all characters $\chi : \mathbb{T} \to \mathbb{C}$ is finite.

2) In general, $\sum_{\chi \in \hat{\mathbb{T}}} V(\chi) \neq V$; i.e. V does not have a basis consisting of $\mathbb{T}$-eigenvectors. (For example, if $\mathbb{T} = \langle T \rangle$, then such a basis exists if and only if T is diagonalizable.) If such a basis exists, then $\mathbb{T}$ is called a *semi-simple* algebra.

3) However, if $\mathbb{T}$ is $*$-*closed* with respect to a hermitian pairing $\langle \, , \, \rangle$ on V, then $\mathbb{T}$ is semi-simple. In other words, if $\mathbb{T}$ has the property that for every operator $T \in \mathbb{T}$, its adjoint T^* is also in $\mathbb{T}$, then

$$V = \bigoplus_{\chi \in \hat{\mathbb{T}}} V(\chi).$$

Here the adjoint $T^* \in \mathrm{End}_{\mathbb{C}}(V)$ of an operator T is defined by

$$\langle T^*(v), w \rangle = \langle v, T(w) \rangle \quad \text{for all } v, w \in V.$$

Proof of Theorem 1.39: We first show that $\mathbf{S}_k$ has a basis consisting of $\mathbb{T}$-eigenfunctions. By Proposition 1.38 we know that $T_n^* = T_n$, for all $n \geq 1$, and hence $\mathbb{T}$ is $*$-closed (as an algebra acting on $\mathbf{S}_k$). Thus, by the above Fact 3) it follows that $V = \mathbf{S}_k$ has a basis consisting of $\mathbb{T}$-eigenfunctions, and hence the same is true for $\mathbf{M}_k = \mathbb{C}E_k \oplus \mathbf{S}_k$ because E_k is a $\mathbb{T}$-eigenfunction by Example 1.34(a). Now if $f_1 = E_k, f_2, \ldots f_r$ is such a basis of $\mathbf{M}_k$, then $a_1(f_i) \neq 0$ by Theorem 1.28, and so we can replace f_i by $\tilde{f}_i = f_i / a_1(f_i)$ to obtain a basis consisting of normalized $\mathbb{T}$-eigenfunctions.

It remains to show that *every* normalized $\mathbb{T}$-eigenfunction f is one of the $\tilde{f}_i$'s. Let $\chi_f : \mathbb{T} \to \mathbb{C}$ denote the associated character of f. Then $\chi_f = \chi_{\tilde{f}_i}$ for some i, and so by the multiplicity 1 result (Corollary 1.30) we have $f = c\tilde{f}_i$, for some $c \in \mathbb{C}$. But since f and $\tilde{f}_i$ are both normalized, it follows that $f = \tilde{f}_i$.

We can now prove the result of Hecke and Petersson which was mentioned at the beginning of this section.

Corollary 1.40 *There are precisely* $\dim \mathbf{M}_k$ *non-zero modular forms* $f \in \mathbf{M}_k$ *whose Fourier coefficients are multiplicative, and these form a basis of* $\mathbf{M}_k$.

Proof. Let $0 \neq f \in \mathbf{M}_k$. By Theorem 1.35, the $a_n(f)$'s are multiplicative if and only if f is a normalized $\mathbb{T}$-eigenfunction. Thus, the assertion follows from Theorem 1.39.

Chapter 2

Modular Forms for Higher Levels

2.1 Introduction

The study of modular forms and functions of higher level was initiated by F. Klein in 1879. His first main motivation for this was to try to understand the Galois group G_N of the so-called *modular equation*

$$\Phi_N(j, j') = 0$$

which is the minimal polynomial of the function $j'(\tau) := j(N\tau)$ over $\mathbb{C}(j)$, where j is usual j-function on $\mathfrak{H}$ and $N \geq 2$ is an integer. Klein had discovered a year earlier that if $N = p$ is prime (and $p \leq 13$), then $G_p \simeq \mathrm{PSL}_2(\mathbb{Z}/p\mathbb{Z})$. He then realized that the functions in the splitting field F_N of Φ_N (over $\mathbb{C}(j)$) are invariant under the action of a (normal) subgroup $\Gamma(N) \leq \mathrm{SL}_2(\mathbb{Z})$, and this led him to study such functions from the point of view of their transformation properties. Klein himself considered this viewpoint as a natural manifestation of his *Erlanger Programm* (of 1872) which proposes that mathematical objects should be classified by their transformation groups.

Later in 1885 Klein showed how similar ideas can be used to study the division values $\wp(\frac{a+b\tau}{N})$ of the Weierstrass $\wp$-function, and this led him to study modular *forms* of higher level. As he explained in some detail, many of the earlier constructions of Jacobi, Legendre, Hermite and many others in the theory of elliptic functions fit into this point of view and have a much more natural interpretation here.

In the meanwhile Klein and his co-workers and students Fricke, Gierster and Hurwitz had undertaken a systematic study of the geometric properties of the Riemann surface $\Gamma(N)\backslash\mathfrak{H}$. Gierster and Hurwitz were particularly interested in applications to number theory such as class number relations of imaginary quadratic fields, a topic that had been first investigated by Kronecker in 1857 via the theory of complex multiplication of elliptic curves. To generalize this, Klein, Gierster and Hurwitz developed a theory of *modular correspondences* of higher level which generalized the modular equation (and which were the basis of Hecke's operators). It is interesting to note in his (successful) attempts to derive these class-number relations, Hurwitz established a general *trace formula* which eventually became the Leftschetz–Eichler–Selberg (etc.) trace formula.

2.2 Basic Definitions and Properties

2.2.1 Congruence subgroups

As before, the group of all integral 2×2 matrices with determinant 1 is called the *modular group* and is denoted by

$$\Gamma(1) = \mathrm{SL}_2(\mathbb{Z}) = \{A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{M}_2(\mathbb{Z}) : \det A = 1\}.$$

For any integer $N > 1$, the *principal congruence subgroup of level* N is

$$\Gamma(N) = \mathrm{Ker}\left(\mathrm{SL}_2(\mathbb{Z}) \stackrel{\mathrm{mod}\ N}{\longrightarrow} \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})\right) = \left\{A \in \Gamma(1) : A \equiv \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \pmod N\right\}.$$

Definition. A subgroup $\Gamma \leq \Gamma(1)$ is called a *congruence subgroup* if it contains $\Gamma(N)$ for some N, i.e. if $\Gamma(N) \leq \Gamma \leq \Gamma(1)$. (The smallest such N is called the *level* of Γ.)

Remark 2.1 (a) Clearly, each congruence subgroup $\Gamma \leq \Gamma(1)$ has finite index in $\Gamma(1)$ (since $\Gamma(N)$ does). Its index, or rather, that of the related group $\pm\Gamma = \Gamma \cup -\Gamma$ is denoted by

$$\mu(\Gamma) = [\Gamma(1) : \pm\Gamma].$$

Note, however, that not every subgroup of finite index is a congruence subgroup. For example, for each odd number $n > 1$, there is a normal subgroup of index $6n^2$ which is not a congruence subgroup (cf. Newman [Ne], p. 150).

(b) If Γ_1 and Γ_2 are congruence subgroups, then so is $\Gamma_1 \cap \Gamma_2$ because

$$\Gamma(N) \cap \Gamma(M) = \Gamma(\mathrm{lcm}(N, M)).$$

(c) If Γ is a congruence subgroup and $\alpha \in \mathrm{GL}_2^+(\mathbb{Q}) := \{g \in \mathrm{GL}_2(\mathbb{Q}) : \det(g) > 0\}$, then $\alpha^{-1}\Gamma\alpha \cap \Gamma(1)$ is also a congruence subgroup. In fact, by multiplying α by a suitable scalar matrix we may assume without loss of generality that $\alpha \in \mathrm{GL}_2^+(\mathbb{Q}) \cap \mathrm{M}_2(\mathbb{Z})$, and then

$$\alpha^{-1}\Gamma(N)\alpha \cap \Gamma(1) \supset \Gamma(ND), \quad \text{where } D = \det(\alpha).$$

In particular, Γ and $\Gamma_1 = \alpha^{-1}\Gamma\alpha$ are *commensurable* subgroups, i.e. $\Gamma \cap \Gamma_1$ has finite index in both Γ and Γ_1.

Example 2.2 The following congruence subgroups are of fundamental importance for much of what follows:

$$\begin{aligned} \Gamma_0(N) &= \left\{A \in \Gamma(1) : A \equiv \left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right) \pmod N\right\}, \\ \Gamma_1(N) &= \left\{A \in \Gamma(1) : A \equiv \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right) \pmod N\right\}. \end{aligned}$$

Note that we have the inclusions $\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \Gamma(1)$, and that $\Gamma(N) \trianglelefteq \Gamma(1)$ and $\Gamma_1(N) \trianglelefteq \Gamma_0(N)$ are normal subgroups with quotients

$$\Gamma(1)/\Gamma(N) \simeq \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}), \quad \Gamma_1(N)/\Gamma(N) \simeq \mathbb{Z}/N\mathbb{Z} \quad \text{and} \quad \Gamma_0(N)/\Gamma_1(N) \simeq (\mathbb{Z}/N\mathbb{Z})^\times.$$

In particular, the respective indices are

$$[\Gamma(1) : \Gamma(N)] \;=\; \#\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}) \;=\; N^3 \prod_{p|N} \left(1 - \frac{1}{p^2}\right),$$

and

$$[\Gamma_0(N) : \Gamma_1(N)] \;=\; \phi(N) \;=\; N \prod_{p|N} \left(1 - \frac{1}{p}\right).$$

From this and the fact that $[\Gamma_1(N) : \Gamma(N)] = N$, we see that

$$[\Gamma(1) : \Gamma_0(N)] = \psi(N) := N \prod_{p|N} \left(1 + \frac{1}{p}\right).$$

Thus, since $-1 \in \Gamma_0(N)$ and $-1 \notin \Gamma_1(N)$ for $N \geq 3$, we obtain

$$\begin{aligned} \mu(\Gamma_0(N)) &= \psi(N) \\ \mu(\Gamma_1(N)) &= \tfrac{1}{2}\phi(N)\psi(N), \quad \text{if } N \geq 3, \\ \mu(\Gamma(N)) &= \tfrac{1}{2}N\phi(N)\psi(N), \quad \text{if } N \geq 3. \end{aligned}$$

Note that we can also write $\Gamma_0(N)$ in the form

$$\Gamma_0(N) \;=\; \alpha_N \Gamma(1) \alpha_N^{-1} \cap \Gamma(1) \;=\; \beta_N^{-1} \Gamma(1) \beta_N \cap \Gamma(1), \tag{2.1}$$

where $\alpha_N := \left(\begin{smallmatrix} 1 & 0 \\ 0 & N \end{smallmatrix}\right)$ and $\beta_N := N\alpha_N^{-1} = \left(\begin{smallmatrix} N & 0 \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{Q})$, for we have

$$\alpha_N \begin{pmatrix} a & b \\ c & d \end{pmatrix} \alpha_N^{-1} = \begin{pmatrix} a & \frac{b}{N} \\ cN & d \end{pmatrix}. \tag{2.2}$$

For later purposes we observe that (2.2) also shows that we have the inclusion

$$\Gamma_1(N) \leq \beta_d^{-1} \Gamma_1(M) \beta_d, \quad \text{if } dM|N. \tag{2.3}$$

Remark. Other natural examples of congruence subgroups are the transposes of the above groups:

$$\Gamma^0(N) = \{g^t : g \in \Gamma_0(N)\} \quad \text{and} \quad \Gamma^1(N) = \{g^t : g \in \Gamma_1(N)\}.$$

Note that $S^{-1}\Gamma_0(N)S = \Gamma^0(N)$ and $S^{-1}\Gamma_1(N)S = \Gamma^1(N)$ where, as before, $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

2.2.2 Modular Functions

Although we will be mainly interested in modular functions on congruence subgroups, it is useful to define them for an arbitrary subgroup $\Gamma \leq \mathrm{GL}_2^+(\mathbb{Q})$. As before, each matrix $g \in GL_2^+(\mathbb{Q})$ acts on the upper half-plane $\mathfrak{H}$ as a fractional linear transformation.

Definition. A *modular function* (of weight 0) on a subgroup $\Gamma \leq \mathrm{GL}^+(\mathbb{Q})$ is function on $\mathfrak{H}$ such that

1) f is meromorphic on $\mathfrak{H}$;

2) $f \circ \gamma = f$, for all γ in Γ;

3) for every $g \in \Gamma(1)$, there is an integer $N = N_g \geq 1$ such that $f \circ g$ has a Puiseaux series expansion in $q_N = e^{2\pi iz/N}$:

$$(f \circ g)(z) = \sum a_{n,g} q_N^n \quad \text{with } a_{n,g} = 0 \text{ for } n << 0. \tag{2.4}$$

In other words, a modular function on Γ is a weakly modular function of weight 0 which satisfies condition 3).

Remark 2.3 (a) The set $\mathcal{M}(\Gamma)$ of all modular functions on Γ is a field containing the field $\mathbb{C}$ of constant functions, as is immediate from the definition. Note that $\mathcal{M}(\pm\Gamma) = \mathcal{M}(\Gamma)$ because -1 acts trivially on $\mathfrak{H}$.

(b) If $\Gamma_1, \Gamma_2 \leq \mathrm{GL}_2^+(\mathbb{Q})$ are any two subgroups, then it is immediate from the definition that

$$\mathcal{M}(\Gamma_1) \cap \mathcal{M}(\Gamma_2) \;=\; \mathcal{M}(\langle \Gamma_1, \Gamma_2 \rangle),$$

where $\langle \Gamma_1, \Gamma_2 \rangle \leq \mathrm{GL}_2^+(\mathbb{Q})$ denotes the subgroup generated by Γ_1 and Γ_2. In particular, if $\Gamma_1 \leq \Gamma_2$ is a subgroup, then $\mathcal{M}(\Gamma_1) \supset \mathcal{M}(\Gamma_2)$ and we have more precisely that

$$\mathcal{M}(\Gamma_2) = \mathcal{M}(\Gamma_1)^{\Gamma_2} := \{f \in \mathcal{M}(\Gamma_1) : f \circ \gamma = f, \forall \gamma \in \Gamma_2\}.$$

(c) For any $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ and $\Gamma \leq \mathrm{GL}^+(\mathbb{Q})$ the map $f \mapsto \alpha^* f := f \circ \alpha$ induces an isomorphism

$$\alpha^* : \mathcal{M}(\Gamma) \xrightarrow{\sim} \mathcal{M}(\alpha^{-1}\Gamma\alpha).$$

[Indeed, if $f \in \mathcal{M}(\Gamma)$ and $\gamma = \alpha^{-1}\gamma_1\alpha \in \alpha^{-1}\Gamma\alpha$, then $(\alpha^* f) \circ \gamma = (f \circ \alpha) \circ (\alpha^{-1}\gamma_1\alpha) = (f \circ \gamma_1) \circ \alpha = f \circ \alpha = \alpha^* f$, so f is weakly modular on $\alpha^{-1}\Gamma\alpha$. Moreover, $\alpha^* f$ also satisfies property 3) by Lemma 2 of [Ko], p. 127, and so $\alpha^* f \in \mathcal{M}(\alpha^{-1}\Gamma\alpha)$. By replacing Γ by $\alpha^{-1}\Gamma\alpha$ and α by α^{-1}, we see that the map α^* is an isomorphism.]

Thus, if $\Gamma \trianglelefteq \Gamma_1$ is a normal subgroup of a subgroup Γ_1, then $g_1^* f \in \mathcal{M}(g_1^{-1}\Gamma g_1) = \mathcal{M}(\Gamma), \forall f \in \mathcal{M}(\Gamma), g_1 \in \Gamma_1$, and hence the quotient group Γ_1/Γ acts as a group of automorphisms of $\mathcal{M}(\Gamma)/\mathcal{M}(\Gamma_1)$. In particular, if $[\Gamma_1 : \Gamma] < \infty$, then it follows from Galois theory that $\mathcal{M}(\Gamma)$ is a finite Galois extension of $\mathcal{M}(\Gamma_1)$.

(d) If Γ is commensurable with $\Gamma(1)$, then the $N = N_g$ appearing in (2.4) can made more precise. For this, let $N_g(\Gamma) := \min\{n : \pm T^n \in g^{-1}\Gamma g\} \leq [\Gamma(1) : \Gamma \cap \Gamma(1)]$. Now if f

is a weakly modular function on Γ, then $(f \circ g) \circ T^N = f \circ g$, where $N = N_g(\Gamma)$, i.e. $f \circ g$ is a periodic function with period N, and hence has a Fourier expansion in q_N. Thus, we see that (2.4) holds for $f \circ g$ for some N if and only if holds for $N = N_g(\Gamma)$.

Note that since the number $N_g(\Gamma)$ and the expansion (2.4) of f only depend on the image $\overline{g(\infty)}$ of $g(\infty)$ in the set $\text{cusps}(\Gamma) = \Gamma\backslash(\mathbb{Q} \cup \{\infty\})$ (as is easy see), it follows that condition 3) has be checked for only finitely many $g \in \Gamma(1)$ because the set $\text{cusps}(\Gamma)$, called the set of *cusps* of Γ, is a finite set. Accordingly we call (2.4) the *Laurent expansion of f at the cusp $\overline{g(\infty)}$*. Moreover, the number $N_g(\Gamma)$ is called the *fan-width* of Γ at the cusp $\overline{g(\infty)}$.

Example 2.4 (a) By Remark 2.3(d) we see that $\mathcal{M}(\Gamma(1)) = \mathbf{A}_0$ (because $N_g(\Gamma(1)) = 1, \forall g \in \Gamma(1)$); i.e. a modular function on $\Gamma(1)$ is the same a modular function of weight 0 in the sense of §1.1. Thus $\mathcal{M}(\Gamma(1)) = \mathbb{C}(j)$ by Theorem 1.2. Moreover, in this case $\Gamma = \Gamma(1)$ acts transitively on $\mathbb{Q} \cup \{\infty\}$, so $\#\text{cusps}(\Gamma(1)) = 1$, i.e. there is only one cusp.

(b) For any $\alpha \in \text{GL}_2^+(\mathbb{Q})$ and $f \in \mathcal{M}(\Gamma)$ we have by Remark 2.3(c) that $f \circ \alpha \in \mathcal{M}(\alpha^{-1}\Gamma\alpha) \subset \mathcal{M}(\Gamma \cap \alpha^{-1}\Gamma\alpha)$. In particular, since $j_N = j \circ \beta_N$, we see that

$$\mathbb{C}(j, j_N) \subset \mathcal{M}(\Gamma(1) \cap \beta_N^{-1}\Gamma(1)\beta_N) = \mathcal{M}(\Gamma_0(N)),$$

where the latter equality follows from (2.1). Thus, j_N is a modular function of higher level, for $j_N \notin \mathcal{M}(\Gamma(1))$, as can seen either directly form the results of chapter 1 or from the facts that $\mathcal{M}(\Gamma_0(N)) = \mathbb{C}(j, j_N)$ and that $\mathcal{M}(\Gamma_0(N)) \neq \mathcal{M}(\Gamma(1))$, which will be established below.

(c) The "division values" of the Weierstrass $\wp$-function give rise to modular functions as follows. For $z \in \mathbb{C}$ and a lattice $L \subset \mathbb{C}$, define the *Weber function* f_0 by

$$f_0(z, L) = -2^7 3^5 \frac{g_2(L)g_3(L)}{\Delta(L)} \wp_L(z),$$

where $\wp_L$ is the Weierstrass $\wp$-function with respect to the lattice L; cf. subsection 1.4.2. Moreover, for $a = (a_1, a_2) \in \mathbb{Q}^2 \setminus \mathbb{Z}^2$ and $\tau \in \mathfrak{H}$ put

$$f_a(\tau) = f_0(a[\tau], L(\tau)) = f_0(a_1\tau + a_2, \mathbb{Z}\tau + \mathbb{Z}).$$

where $[\tau]$ denotes the column vector $[\tau] = (\tau, 1)^t$ and where a is viewed as a row vector. The functions f_a are called *Fricke functions* and satisfy the transformation law

$$f_{ag}(\tau) = f_a(g(\tau)), \quad \forall g \in \Gamma(1). \tag{2.5}$$

To see this, note first that the Weber function f_0 satisfies the homogeneity property

$$f_0(cz, cL) = f_0(z, L), \quad \forall z, c \in \mathbb{C}, c \neq 0,$$

which follows immediately from the fact that $h(L) := g_2(L)g_3(L)/\Delta(L)$ is a lattice function of weight $-2 = (4 + 6 - 12)$ (so $h(cL) = c^2 h(L)$) and from the fact that

$\wp_{cL}(cz) = c^{-2}\wp_L(z)$ (which is clear from the definition of $\wp_L$). Thus, since $(ag)[\tau] = a[g(\tau)]j(g,\tau)$ and $j(g,\tau)L(g(\tau)) = L(\tau)$, we obtain $f_a(g(\tau)) = f_0(a[g(\tau)], L(g(\tau))) = f_0(a[g(\tau)]j(g,\tau), j(g,\tau)L(g(\tau)) = f_0((ag)[\tau], L(\tau))f_{ag}(\tau)$, which proves the relation (2.5).

Next we note that

$$f_a = f_{a'} \;\Leftrightarrow\; a \equiv \pm a' \pmod{\mathbb{Z}^2}$$

because $\wp_L(z_1) = \wp_L(z_2) \Leftrightarrow z_1 \equiv \pm z_2 \pmod{L}$ by Proposition 1.19. Thus, if we write $f_{r,s,N} = f_{(r/N,s/N)}$ when $r, s, N \in \mathbb{Z}$, $N \geq 1$ and $(r,s) \not\equiv (0,0) \pmod{N}$, then we have

$$f_{r,s,N} = f_{r',s',N} \;\Leftrightarrow\; (r,s) \equiv \pm(r',s') \pmod{N}. \tag{2.6}$$

In particular, since $(r,s)g \equiv (r,s) \pmod{N}$ when $g \in \Gamma(N)$, we see that $f_{r,s,N} \circ g = f_{r,s,N}$, for all $g \in \Gamma(N)$.

To see that $f_{r,s,N}$ is holomorphic on $\mathfrak{H}$ and is meromorphic at the cusps, we will use the following expansion of the Weierstrass function $\wp$-function $\wp(z,\tau) = \wp_{L(\tau)}(z)$ in terms of $q = e^{2\pi i\tau}$ and $w = e^{2\pi iz}$ which is valid for $|q| < |w| < |q|^{-1}$:

$$\frac{12}{(2\pi i)^2}\wp(z,\tau) = 1 + \frac{12w}{(1-w)^2} + 12\sum_{m,n=1}^{\infty} nq^{mn}(w^n + w^{-n} - 2)$$

(This formula is easily established by considering a suitable expansion of the Weierstrass $\wp$-function; cf. Lang[La0], p. 46 for details.)

Substituting $z = \frac{r\tau+s}{N}$ yields $w = q_N^r\zeta_N^s$, where $\zeta_n = e^{2\pi i/N} \in \mathbb{C}$, and so it is clear that $\wp_{r,s,N}(\tau) := \wp\left(\frac{r\tau+s}{N}, \tau\right)$ has a power series expansion in q_N which converges everywhere on $\mathfrak{H}$. (Here we assume that $0 \leq r < N$, which is no restriction by (2.6).) Thus, since $h(\tau) = h(L(\tau)) \in \mathbf{A}_{-2}$, it follows that $f_{r,s,N}(\tau) = ch(\tau)\wp_{r,s,N}(\tau)$ (where $c \in \mathbb{C}$) is holomorphic on $\mathfrak{H}$ and has a Laurent expansion in q_N. Moreover, since the $f_{r,s,N}$'s are permuted by the action of $\Gamma(1)$ (when N is fixed), it follows that $f_{r,s,N}$ is meromorphic at the cusps, and so $f_{r,s,N} \in \mathcal{M}(\Gamma(N))$.

We now prove the following fundamental result.

Theorem 2.5 *For each $N \geq 1$, the field $F_N := \mathcal{M}(\Gamma(N))$ of modular functions of level N is generated by j and the two Fricke functions $f_{1,0,N}$ and $f_{0,1,N}$. Thus*

$$F_N = \mathbb{C}(j, f_{1,0,N}, f_{0,1,N}) = \mathbb{C}(j, \{f_{r,s,N}\}_{r,s}).$$

Moreover, F_N is a Galois extension of $F_1 = \mathbb{C}(j)$ with Galois group

$$\mathrm{Gal}(F_N/F_1) \simeq \Gamma(1)/\pm\Gamma(N) \simeq SL_2(\mathbb{Z}/N\mathbb{Z})/(\pm 1).$$

Proof. By Example 2.4(c) we know that $f_{r,s,N} \in F_N$, and so $F := \mathbb{C}(j, f_{1,0,N}, f_{0,1,N}) \subset \mathbb{C}(j, \{f_{r,s,N}\}) \subset F_N$. Thus, the first assertion follows once we have shown that $F = F_N$.

Now by Remark 2.3(c) we know that $F_N/F_1 = \mathbb{C}(j)$ is a Galois extension with group $\Gamma(1)/K$, where $K = \{g \in \Gamma(1) : f \circ g = f, \forall f \in F_N\}$. Let $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma(1)$. Since

$(1,0)g = (a,b)$ we have by (2.5) that $f_{1,0,N} \circ g = f_{a,b,N}$ and similarly $f_{0,1,N} \circ g = f_{c,d,N}$. Thus, if $g \in K$ then $f_{1,0,N} = f_{a,b,N}$ and $f_{0,1,N} = f_{c,d,N}$, and so by (2.6) we have $(a,b) \equiv \pm(1,0) \pmod N$ and $(c,d) \equiv \pm(1,0) \pmod N$, i.e., $g \in \pm\Gamma(N)$. Thus $\mathrm{Gal}(F_N/F_1) = \Gamma(1)/(\pm\Gamma(N))$. In addition, we see that $\mathrm{Gal}(F_N/F) = \pm\Gamma(N)/(\pm(\Gamma(N)) = 1$, and so $F = F_N$.

Corollary 2.6 *If $\Gamma \leq \Gamma(1)$ is a congruence subgroup of level N, then*

$$\mathrm{Gal}(F_N/\mathcal{M}(\Gamma)) = (\pm\Gamma)/(\pm\Gamma(N))$$

In particular, $\mathcal{M}(\Gamma)$ is a finite extension of $F_1 = \mathbb{C}(j)$ of degree $\mu(\Gamma)$, i.e. we have $[\mathcal{M}(\Gamma) : \mathbb{C}(j)] = \mu(\Gamma)$. Moreover, every intermediate field $F_1 \subset F \subset F_N$ is of the form $F = \mathcal{M}(\Gamma)$ for some congruence subgroup Γ.

Proof. Since $(\pm\Gamma)/(\pm\Gamma(N)) \leq \Gamma(1)/(\pm\Gamma(N)) = \mathrm{Gal}(F_N/F_1)$ and since $\mathcal{M}(\Gamma) = F_N^\Gamma$ by Remark 2.3(b), the first assertion is clear by Galois theory. Thus $[F_N : \mathcal{M}(\Gamma)] = [\pm\Gamma : \pm\Gamma(N)]$, and hence $[\mathcal{M}(\Gamma) : F_1] = [G(1) : \pm\Gamma] = \mu(\Gamma)$.

Finally, since F_N/F_1 is Galois with group $G = \Gamma(1)/(\pm\Gamma(N))$, we have by Galois theory that $F = (F_N)^H$, for some subgroup $H = \Gamma/(\pm\Gamma(N)) \leq G$, and so $F = \mathcal{M}(\Gamma)$ by Remark 2.3(b).

Corollary 2.7 *For any $N \geq 2$ we have*

$$\mathcal{M}(\Gamma_1(N)) = \mathbb{C}(j, f_{0,1,N}) \quad \textit{and} \quad \mathcal{M}(\Gamma^1(N)) = \mathbb{C}(j, f_{1,0,N}).$$

Proof. If $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma(1)$, then $f_{0,1,N} \circ g = f_{c,d,N}$ (cf. the proof of Theorem 2.5) and so by (2.6) we have that $f_{0,1,N} \circ g = f_{0,1,N}$ if and only if $(c,d) \equiv \pm(0,1) \pmod N$, i.e. if and only if $g \in \pm\Gamma_1(N)$ (because $\det(g) \equiv 1 \pmod N$). This means that $\mathrm{Gal}(F_N/\mathbb{C}(j, f_{0,1,N})) = \pm\Gamma_1(N)/(\pm\Gamma(N))$, and so by Galois theory and Corollary 2.6 it follows that it $\mathcal{M}(\Gamma_1(N)) = \mathbb{C}(j, f_{0,1,N})$, as asserted. The proof of the second equation is analogous.

In order to understand other fields of modular functions, it is useful to have information about the following group $G(\mathcal{M}(\Gamma))$.

Notation. For any set $\mathcal{M}$ of meromorphic functions on $\mathfrak{H}$, let

$$G(\mathcal{M}) = \{\alpha \in \mathrm{GL}_2^+(\mathbb{Q}) : f \circ \alpha = f,\ \forall f \in \mathcal{M}\}.$$

Clearly, $G(\mathcal{M})$ is a subgroup of $\mathrm{GL}_2^+(\mathbb{Q})$ and if $\mathcal{M}_1 \subset \mathcal{M}_2$, then $G(\mathcal{M}_1) \geq G(\mathcal{M}_2)$. Moreover, for any $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ we have

$$G(\alpha^*\mathcal{M}) = \alpha^{-1}G(\mathcal{M})\alpha, \tag{2.7}$$

for $\beta \in \mathrm{GL}_2^+(\mathbb{Q})$, then $\beta \in G(\alpha^*\mathcal{M}) \Leftrightarrow (f \circ \alpha) \circ \beta = f \circ \alpha, \forall f \in \mathcal{M} \Leftrightarrow f \circ (\alpha\beta\alpha^{-1}) = f, \forall f \in \mathcal{M} \Leftrightarrow \alpha\beta\alpha^{-1} \in G(\mathcal{M}) \Leftrightarrow \beta \in \alpha^{-1}G(\mathcal{M})\alpha$.

We now prove the following refinement of Corollary 2.6:

Theorem 2.8 *If $\Gamma \leq \Gamma(1)$ is any congruence subgroup, then*

$$G(\mathcal{M}(\Gamma)) = \mathbb{Q}^\times \Gamma. \tag{2.8}$$

To prove this, we shall use the following simple lemma which is in fact a special case of the so-called *Smith normal form* for integral matrices; cf. Newman[Ne], p. 26.

Lemma 2.9 *If $g \in \mathrm{GL}_2^+(\mathbb{Q}) \cap M_2(\mathbb{Z})$, then there exist $g_1, g_2 \in \mathrm{SL}_2(\mathbb{Z})$ and $m, n \in \mathbb{Z}^+$ such that $g = m g_1 \beta_n g_2$.*

Proof. Write $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, and put $m = \gcd(a, b, c, d)$. Then $g' = \frac{1}{m} g \in \mathrm{GL}_2^+(\mathbb{Q}) \cap M_2(\mathbb{Z})$ is a primitive matrix of determinant $n = \det(g)/m^2$, and so by [Sch], p. 135 (or [Ne], p. 26) there exist $g_1, g_2 \in \Gamma(1)$ such that $g' = g_1 \beta_n g_2$.

Proof of Theorem 2.8. First note that it follows from the definitions that we have $\Gamma \leq G(\mathcal{M}(\Gamma))$. Moreover, since scalar matrices act as the identity on $\mathfrak{H}$, it is clear that $\mathbb{Q}^\times \leq G(\mathcal{M}(\Gamma))$. Thus, since $G(\mathcal{M}(\Gamma))$ is a subgroup of $\mathrm{GL}_1^+(\mathbb{Q})$, we see that $\mathbb{Q}^\times \Gamma \leq G(\mathcal{M}(\Gamma))$.

To prove the opposite inclusion, we first consider the case $\Gamma = \Gamma(1)$. Suppose there exists $\alpha \in G(\mathcal{M}(\Gamma(1))) \setminus \mathbb{Q}^\times \Gamma(1)$. Then by replacing α by $c\alpha$, we may assume that $\alpha \in \mathrm{GL}_2^+(\mathbb{Q}) \cap M_2(\mathbb{Z})$. Thus, by the above Lemma 2.9 we have that $\beta_n \in G(\mathcal{M}(\Gamma(1)))$, for some $n > 1$, and so $j \circ \beta_n = j$ because $j \in \mathcal{M}(\Gamma(1))$. This implies that $j(2ni) = j(\beta_n(2i)) = j(2i)$. But since both $2ni$ and $2i$ lie in the fundamental domain D of $\mathrm{SL}_2(\mathbb{Z})$) (cf. the proof of Proposition 1.3), this contradicts Proposition 1.21. Thus, no such α exists, and so $G(\mathcal{M}(\Gamma(1))) = \mathbb{Q}^\times \Gamma(1)$.

Now suppose that Γ is any congruence subgroup. Then $\Gamma(1) \geq \Gamma \geq \Gamma(N)$ for some N. Let $\alpha \in G(\mathcal{M}(\Gamma))$. Since $G(\mathcal{M}(\Gamma)) \leq G(\mathcal{M}(\Gamma(1))) = \mathbb{Q}^\times \Gamma(1)$ (by what was just proved), we see that $\alpha = cg$ with $c \in \mathbb{Q}^\times$ and $g \in \Gamma(1)$. Then $g \in G(\mathcal{M}(\Gamma)) \cap \Gamma(1)$, and so the image of g in $\mathrm{Gal}(F_N/F_1)$ actually lies in $\mathrm{Gal}(F_N/\mathcal{M}(\Gamma)) = \pm\Gamma/(\pm\Gamma(N))$, the latter by Corollary 2.6. Thus $g \in \pm\Gamma$, i.e. $\alpha \in \mathbb{Q}^\times \Gamma$. This proves $G(\mathcal{M}(\Gamma)) \leq \mathbb{Q}^\times \Gamma$, and so the theorem follows.

It is useful to generalize this result to *generalized congruence subgroups* which are defined as follows.

Definition. A subgroup $\Gamma \leq \mathrm{GL}_2^+(\mathbb{Q})$ is called a *generalized congruence subgroup* if $\alpha \Gamma \alpha^{-1}$ is a congruence subgroup for some $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$.

Remark 2.10 Let $\mathcal{C} = \{\Gamma : \Gamma(1) \geq \Gamma \geq \Gamma(N), \text{for some } N\}$ denote the set of congruence subgroups, and $\mathcal{G} = \cup_\alpha \alpha^{-1} \mathcal{C} \alpha$ the set of all generalized congruence subgroups. Then:

(1) $\Gamma \in \mathcal{G} \Rightarrow \alpha^{-1} \Gamma \alpha \in \mathcal{G}, \forall \alpha \in \mathrm{GL}_2^+(\mathbb{Q})$;

(2) $\Gamma \in \mathcal{G} \Rightarrow \mathrm{SL}_2(\mathbb{Q}) \geq \Gamma \geq \Gamma(N)$, for some N;

(3) $\Gamma_1 \in \mathcal{C}, \Gamma_2 \in \mathcal{G} \Rightarrow \Gamma_1 \cap \Gamma_2 \in \mathcal{C}$;

(4) $\Gamma_1, \Gamma_2 \in \mathcal{G} \Rightarrow \Gamma_1 \cap \Gamma_2 \in \mathcal{G}$.

[Indeed, (1) is clear. For (2) we note that if $\Gamma = \alpha^{-1}\Gamma_1\alpha$ with $\Gamma_1 \in \mathcal{C}$, then $\det(\alpha^{-1}g_1\alpha) = \det(g_1) = 1$, $\forall g_1 \in \Gamma_1$, and so $\Gamma \leq \mathrm{SL}_2(\mathbb{Q})$. Moreover, since $\Gamma_1 \geq \Gamma(N_1)$ for some N_1, we have $\Gamma = \alpha^{-1}\Gamma_1\alpha \geq \alpha^{-1}\Gamma(N_1)\alpha \geq \Gamma(N_1D)$, for some D by Remark 2.1(c), which proves (2). To prove (3), we use the fact that $\Gamma_i \geq \Gamma(N_i)$ by (2). Thus, $\Gamma(1) \geq \Gamma_1 \geq \Gamma_1 \cap \Gamma_2 \geq \Gamma(N_1) \cap \Gamma(N_2) \geq \Gamma(N_1N_2)$ by Remark 2.1(b), and so $\Gamma_1 \cap \Gamma_2 \in \mathcal{C}$. Finally (4) follows from (3) because if $\alpha^{-1}\Gamma_1\alpha \in \mathcal{C}$, then $\alpha^{-1}(\Gamma_1 \cap \Gamma_2)\alpha = (\alpha^{-1}\Gamma_1\alpha) \cap (\alpha^{-1}\Gamma_2\alpha) \in \mathcal{C}$ by (3), so $\Gamma_1 \cap \Gamma_2 \in \mathcal{G}$.]

Corollary 2.11 *If $\Gamma \leq \mathrm{GL}_2^+(\mathbb{Q})$ is any generalized congruence subgroup, then* (2.8) *holds for Γ, i.e. $G(\mathcal{M}(\Gamma)) = \mathbb{Q}^\times\Gamma$.*

Proof. Put $\Gamma_1 = \alpha\Gamma\alpha^{-1}$. Then $G(\mathcal{M}(\Gamma_1)) = \mathbb{Q}^\times\Gamma_1$ by Theorem 2.8. Thus by Remark 2.3(c) and (2.7) we have $G(\mathcal{M}(\Gamma)) = G(\mathcal{M}(\alpha^{-1}\Gamma_1\alpha)) = C(\alpha^*\mathcal{M}(\Gamma_1)) = \alpha^{-1}G(\mathcal{M}(G_1))\alpha = \alpha^{-1}(\mathbb{Q}^\times\Gamma_1)\alpha = \mathbb{Q}^\times\Gamma$.

Corollary 2.12 *If $\Gamma_1, \Gamma_2 \leq \mathrm{GL}_2^+(\mathbb{Q})$ are two generalized congruence subgroups, then*

$$\mathcal{M}(\Gamma_1) \subset \mathcal{M}(\Gamma_2) \Leftrightarrow \mathbb{Q}^\times\Gamma_1 \geq \mathbb{Q}^\times\Gamma_2 \Leftrightarrow \pm\Gamma_1 \geq \pm\Gamma_2, \tag{2.9}$$

and so $\mathcal{M}(\Gamma_1) = \mathcal{M}(\Gamma_2) \Leftrightarrow \pm\Gamma_1 = \pm\Gamma_2$. Thus

$$\mathcal{M}(\Gamma_1 \cap \Gamma_2) = \mathcal{M}(\Gamma_1)\mathcal{M}(\Gamma_2). \tag{2.10}$$

Proof. If $\mathbb{Q}^\times\Gamma_1 \geq \mathbb{Q}^\times\Gamma_2$, then clearly $\mathcal{M}(\Gamma_1) \subset \mathcal{M}(\Gamma_2)$ (cf. Remark 2.3(b)). Conversely, if $\mathcal{M}(\Gamma_1) \subset \mathcal{M}(\Gamma_2)$, then $G(\mathcal{M}(\Gamma_1)) \geq G(\mathcal{M}(\Gamma_2))$, and so $\mathbb{Q}^\times\Gamma_1 \geq \mathbb{Q}^\times\Gamma_2$ by Corollary 2.11. This proves the first equivalence of (2.9). To prove the second, suppose $g_2 \in \Gamma_2 \leq \mathbb{Q}^\times\Gamma_1$. Then $g_2 = cg_1$ with $c \in \mathbb{Q}^\times$, $g_1 \in \Gamma_1$, so $1 = \det(g_2) = \det(cg_1) = c^2$. Thus $c = \pm 1$, and hence $g_2 \in \pm\Gamma_1$. This proves the second equivalence of (2.9).

To prove (2.10), we first show that $F := \mathcal{M}(\Gamma_1)\mathcal{M}(\Gamma_2) = \mathcal{M}(\Gamma_3)$ for some generalized congruence subgroup Γ_3. Now since $\alpha^{-1}\Gamma_1\alpha$ is a congruence subgroup, then so is $\alpha^{-1}(\Gamma_1 \cap \Gamma_2)\alpha$ (cf. Remark 2.10), and hence $F_1 \subset \alpha^*\mathcal{M}(\Gamma_1) = \mathcal{M}(\alpha^{-1}\Gamma_1\alpha) \subset \alpha^*F = \alpha^*(\mathcal{M}(\Gamma_1)\mathcal{M}(\Gamma_2)) \subset \alpha^*\mathcal{M}(\Gamma_1 \cap \Gamma_2) \subset F_N$ for some N. (Note that since $\Gamma_i \geq \Gamma_1 \cap \Gamma_2$, for $i = 1, 2$, we have $\mathcal{M}(\Gamma_i) \subset \mathcal{M}(\Gamma_1 \cap \Gamma_2)$ and so $\alpha^*F = \alpha^*(\mathcal{M}(\Gamma_1)\mathcal{M}(\Gamma_2)) \subset \alpha^*\mathcal{M}(\Gamma_1 \cap \Gamma_2)$.) Thus, by Corollary 2.6 we have $\alpha^*F = \mathcal{M}(\Gamma)$ for some congruence subgroup Γ, and hence $F = \mathcal{M}(\alpha\Gamma\alpha^{-1}) = \mathcal{M}(\Gamma_3)$.

Now since $G(\mathcal{M}(\Gamma_i)) = \mathbb{Q}^\times\Gamma_i$ by Corollary 2.11, and since it is clear that $G(F) = G(\mathcal{M}(\Gamma_1)\mathcal{M}(\Gamma_2)) = G(\mathcal{M}(\Gamma_1)) \cap G(\mathcal{M}(\Gamma_2))$, we see that $G(\mathcal{M}(\Gamma_3)) = G(F) = \mathbb{Q}^\times(\Gamma_1 \cap \Gamma_2) = G(\mathcal{M}(\Gamma_1 \cap \Gamma_2))$, and so by (2.9) we have $F = \mathcal{M}(\Gamma_3) = \mathcal{M}(\Gamma_1 \cap \Gamma_2)$, which is (2.10).

Remark 2.13 The above results can be viewed as giving a (generalized) Galois correspondence between certain subfields of the *field $F_\infty = \cup F_N$ of all modular functions*, and the set $\mathcal{G}^\pm \subset \mathcal{G}$ consisting of all generalized congruence subgroups $\Gamma \in \mathcal{G}$ with $\pm 1 \in \Gamma$. More precisely, if we let $\mathcal{F}$ denote the set of all *generalized congruence subfields*, i.e. the

set of all subfields $F \subset F_\infty$ with the property that $F \supset \mathbb{C}(j \circ \alpha)$ and $[F : \mathbb{C}(j \circ \alpha)] < \infty$, for some $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, then the above results show that the maps $\Gamma \mapsto \mathcal{M}(\Gamma)$ and $F \mapsto G(F) \cap \mathrm{SL}_2(\mathbb{Q})$ are inverses of each other and hence induce a bijection (Galois correspondence)

$$\mathcal{G}^{\pm} \xrightarrow{\sim} \mathcal{F}$$

between the set $\mathcal{G}^{\pm}$ of generalized congruence subgroups and the set $\mathcal{F}$ of generalized congruence subfields of F_∞.

Example 2.14 If Γ is a congruence subgroup and $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, then it follows from (2.10) that

$$\mathcal{M}(\Gamma \cap \alpha^{-1}\Gamma\alpha) = \mathcal{M}(\Gamma)\alpha^*\mathcal{M}(\Gamma).$$

In particular, since $X_0(N) = \Gamma(1) \cap \beta_N^{-1}\Gamma(1)\beta_N$ (cf. (2.1)), we see that

$$\mathcal{M}(X_0(N)) = \mathbb{C}(j, j_N). \tag{2.11}$$

Note: In [Sh] it is asserted that this follows immediately from Galois theory (i.e. from Corollary 2.6), but there seems to be a gap in the argument.

By elementary field theory, we thus see from (2.11) (together with Corollary 2.6 and Example 2.2) that j_N satisfies a unique monic irreducible polynomial $\Phi_N(x)$ of

$$\deg \Phi_N = [\mathcal{M}(X_0(N) : \mathbb{C}(j)] = \psi(N).$$

with coefficients in $\mathbb{Q}(j)$; this polynomial called the *modular polynomial of order N*. To find an explicit expression for Φ_N, we first introduce the following notation.

Notation. For $N \geq 1$, let $\mathcal{P}_N \subset M_2(\mathbb{Z})$ denote the set of *primitive integral matrices* of determinant n, i.e.

$$\mathcal{P}_N = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in M_2(\mathbb{Z}) : ad - bc = N \text{ and } \gcd(a, b, c, d) = 1\}.$$

Proposition 2.15 *The modular polynomial Φ_N can be written in the form*

$$\Phi_N(x) = \prod_{g \in \Gamma(1)\backslash\mathcal{P}_N} (x - j \circ g) = \prod_{\substack{ad = N \\ 0 \leq b < d \\ gcd(a,b,d) = 1}} \left(x - j \circ \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)\right).$$

To prove this, we shall use the following result which is a refinement of Lemma 2.9 and which is a special case of a general result due to Hermite (cf. [Ne], p. 15) about integral matrices:

Lemma 2.16 *For any $N \geq 1$ we have*

$$\mathcal{P}_N = \Gamma(1)\beta_N\Gamma(1) = \Gamma(1)\alpha_N\Gamma(1) = \dot{\bigcup_{\substack{ad = N \\ 0 \leq b < d \\ gcd(a,b,d) = 1}}} \Gamma(1) \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}. \tag{2.12}$$

Proof. Clearly, if $\alpha \in \mathcal{P}_N$, then $g_1 \alpha g_2 \in \mathcal{P}_N$, for all $g_1, g_2 \in \Gamma(1)$, i.e. $\Gamma(\alpha\Gamma(1) \subset \mathcal{P}_N$. Thus, $\mathcal{P}_N$ contains both double cosets. On the other hand, by Lemma 2.9 we see that $\mathcal{P}_N \subset \Gamma(1)\beta_N\Gamma(1)$, and so we have equality. This proves the first two equalities. For the third (which is Hermite's result), see [Sch], p. 133 or [Ne], p. 15.

Proof of Proposition 2.15. It immediate that the product $\prod_{g\in\Gamma(1)\backslash\mathcal{P}_N}(x - j\circ g)$ does not depend on the choice of the system of representatives $\{g\}$ of the coset space $\Gamma(1)\backslash\mathcal{P}_N$. Thus, the second identity follows directly from Lemma 2.16.

To prove the first, we observe that by Galois theory and Corollary 2.6 we have

$$\Phi_N(x) = \prod(x - j_N \circ g_i) = \prod(x - j \circ \beta_N g_i),$$

where $\{g_i\}$ is a system of representatives of $\Gamma_0(N)\backslash\Gamma(1)$, i.e. $\Gamma(1) = \dot\cup \Gamma_0(N)g_i$. Since $\Gamma_0(N) = \Gamma(1) \cap \beta_N^{-1}\Gamma(1)\beta_N$ (cf. equation (2.1)), it follows that $\{\beta_N g_i\}$ is a system of representatives of $\Gamma(1)\backslash\Gamma(1)\beta_N\Gamma(1)$, and so the first identity follows.

Remark 2.17 (a) As we shall see in more detail below, there is a close connection between modular polynomials and Hecke operators. For example, if N is *squarefree*, then the trace of j_N (with respect to the field extension $\mathcal{M}(X_0(N))/\mathbb{C}(j)$) is essentially the Hecke operator applied to j; explicitly, we have

$$\mathrm{tr}_{\mathcal{M}(X_0(N))/\mathbb{C}(j)}(j_N) = NT_N(j),$$

for by Proposition 2.15 and equation (1.63) we have $\mathrm{tr}(j_N) = \sum j \circ \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) = NT_N(j)$, the latter because the condition $\gcd(a, b, d) = 1$ holds automatically when N is squarefree.

(b) It is easy to see that the coefficients of $\Phi_N(x)$ are polynomials in j; i.e. $\Phi_N(x) \in \mathbb{C}[j][x] = \mathbb{C}[j, x]$. Thus we can write

$$\Phi_N(x) = P_N(x, j) \quad \text{with } P_N \in \mathbb{C}[x, y].$$

In fact, it turns out that the coefficients of P_N are integers (so $P_N \in \mathbb{Z}[x, y]$) which grow very rapidly with N. Furthermore, P_N is symmetric in x and y, i.e. $P_N(y, x) = P_N(x, y)$; cf. [Sch], p. 143-144. Further properties are discussed in Weber[We] III, p. 239-245.

(c) The Galois group of (the splitting field of) Φ_N is:

$$\mathrm{Gal}(\Phi_N) \simeq \mathrm{PSL}_2(\mathbb{Z}/N\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})/Z(\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}));$$

cf. [Sch], p. 148. Since $Z(\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})) = \{cI : c^2 \equiv 1 \pmod N\}$, we see that F_N is the splitting field of Φ_N if $N = p^r$ or $N = 2p^r$ (where p is an odd prime), but in general the splitting field of Φ_N is a proper subfield of F_N.

(d) The roots of the polynomial $\Phi_N(X, X)$ are called the *singular values* of the j-function; in the context of elliptic curves these correspond to CM-elliptic curves (i.e. elliptic curves with complex multiplication). See Lang[La0], p. 143–147, Weber[We], p. 419–423 or Shimura[Sh], p. 109 for more details.

Remark 2.18 Recall that a *Riemann surface* is a connected topological space which is covered by a family of open sets isomorphic to $\mathbb{C}$ with the property that the transition functions are holomorphic functions; cf. Springer[Sp] or Forster[Fo]. For example, the complex plane $\mathbb{C}$, the Riemann sphere $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$ and an elliptic curve $E_L = \mathbb{C}/L$ are all examples of Riemann surfaces.

If Γ is congruence subgroup, then it is easy to see that the quotient space $X'_\Gamma := \Gamma\backslash\mathfrak{H}$ can be made into Riemann surface such that the quotient map $p_\Gamma : \mathfrak{H} \to \Gamma\backslash\mathfrak{H}$ is a holomorphic map; cf. [Sh], p. 17 or [Mi], p. 24. Now since each modular function $f \in \mathcal{M}(\Gamma)$ defines a unique function

$$f_\Gamma : \Gamma\backslash\mathfrak{H} \to \mathbb{C} \cup \{\infty\},$$

such that $f = p_\Gamma^* f_\Gamma := f_\Gamma \circ p_\Gamma$, it follows from conditions 1) and 2) in the definition of a modular function that we have an inclusion $\mathcal{M}(\Gamma) \subset p_\Gamma^*\mathcal{M}(\Gamma\backslash\mathfrak{H})$, where $\mathcal{M}(X'_\Gamma)$ denotes the field of meromorphic functions on the Riemann surface $X'_\Gamma = \Gamma\backslash\mathfrak{H}$.

In order to be able to translate condition 3) into a condition in complex analysis, we first compactify $\mathfrak{H}$ by adding its "cusps":

$$\mathfrak{H}^* := \mathfrak{H} \cup \mathbb{Q} \cup \{\infty\} = \mathfrak{H} \cup \mathbb{P}^1(\mathbb{Q}).$$

Note that the action of $\mathrm{GL}_2^+(\mathbb{Q})$ (and hence of $\Gamma(1)$) on $\mathfrak{H}$ extends naturally to one on $\mathfrak{H}^*$ if we set $\gamma(\infty) = \frac{a}{c}$ for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$.

Now the quotient $X_\Gamma := \Gamma\backslash\mathfrak{H}^*$ can be given the structure of a *compact Riemann surface* which contains $X'_\Gamma := \Gamma\backslash\mathfrak{H}$ as an open subsurface with a finite complement

$$\mathrm{cusps}(\Gamma) = \mathrm{cusps}(X_\Gamma) := X_\Gamma \setminus X'_\Gamma = \Gamma\backslash\mathbb{P}^1(\mathbb{Q})$$

(cf. [Mi], p. 24ff or [Sh], p. 17ff), and then we have

$$\mathcal{M}(\Gamma) = p_\Gamma^*\mathcal{M}(X_\Gamma). \tag{2.13}$$

For example, if $\Gamma = \Gamma(1)$, then $\mathrm{cusps}(\Gamma) = \{P_\infty\}$ consists of one point $P_\infty = p_{\Gamma(1)}(\infty)$, and j defines a isomorphism $j : X(1) := X_{\Gamma(1)} \xrightarrow{\sim} \mathbb{C}_\infty := \mathbb{C} \cup \infty$ such that $j(P_\infty) = \infty$; cf. Proposition 1.21. In particular, $\mathcal{M}(\Gamma(1)) \simeq \mathbb{C}(j)$.

If $\Gamma_1 \leq \Gamma_2$ are two congruence subgroups, then the inclusion map induces a quotient map

$$p_{\Gamma_1,\Gamma_2} : X_{\Gamma_1} = \Gamma_1\backslash\mathfrak{H}^* \to X_{\Gamma_2} = \Gamma_2\backslash\mathfrak{H}^*$$

which is a holomorphic map (of compact Riemann surfaces) of degree

$$\deg(p_{\Gamma_1,\Gamma_2}) = [\pm\Gamma_2 : \pm\Gamma_1] = \mu(\Gamma_1)/\mu(\Gamma_2).$$

Note that since $p_{\Gamma_1,\Gamma_2} \circ p_{\Gamma_1} = p_{\Gamma_2}$, the map $p_{\Gamma_1}^*$ of (2.13) naturally identifies the subfield $p_{\Gamma_1,\Gamma_2}^*\mathcal{M}(X_{\Gamma_2}) \subset \mathcal{M}(X_{\Gamma_1})$ with the subfield $\mathcal{M}(\Gamma_2) \subset \mathcal{M}(\Gamma_1)$. Thus we have

$$\deg(p_{\Gamma_1,\Gamma_2}) = [\mathcal{M}(X_{\Gamma_1}) : p_{\Gamma_1,\Gamma_2}^*\mathcal{M}(X_{\Gamma_2})] = [\mathcal{M}(\Gamma_1) : \mathcal{M}(\Gamma_2)] = \mu(\Gamma_1)/\mu(\Gamma_2).$$

More generally, if $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ is such that $\alpha\Gamma_1\alpha^{-1} \leq \Gamma_2$, then there is a unique holomorphic map $p_{\Gamma_1,\Gamma_2,\alpha} : X_{\Gamma_1} \to X_{\Gamma_2}$ such that

$$p_{\Gamma_2} \circ \alpha = p_{\Gamma_1,\Gamma_2,\alpha} \circ p_{\Gamma_1}. \tag{2.14}$$

2.2.3 Modular Forms

We now define modular forms for an arbitrary subgroup $\Gamma \leq \mathrm{SL}_2(\mathbb{Q})$.

Definition. A *modular form of weight k on Γ* is a map $f : \mathfrak{H} \to \mathbb{C}$ such that

1) f is holomorphic on $\mathfrak{H}$;

2) $f|_k\gamma = f$, for all $\gamma \in \Gamma$.

3) For each $g \in \Gamma(1)$, there is an integer $N = N_g$ such that $f \circ g$ has a Puiseaux series expansion in $q_N = e^{2\pi iz/N}$ with non-negative terms:

$$(f \circ g)(z) \ = \ \sum_{n=0}^{\infty} a_{n,g} q_N^n.$$

Furthermore, a modular form f is called a *cusp form* if we have $a_{0,g} = 0$ for all $g \in \Gamma(1)$.

Note that the set $M_k(\Gamma)$ of all modular forms of weight k on Γ is a $\mathbb{C}$-vector space which contains the set $S_k(\Gamma)$ of all cusp forms as a subspace.

Remark 2.19 (a) More generally, we can also define *automorphic functions* (or *modular functions*) $f \in A_k(\Gamma)$ of weight k on Γ: these are weakly meromorphic functions of weight k on Γ (cf. §1.1) which have a Laurent expansion (2.4) in q_N at each cusp $z = g(\infty) \in \mathbb{Q} \cup \{\infty\}$. Thus, we have

$$f \in M_k(\Gamma) \ \Leftrightarrow \ f \in \mathbb{A}_k(\Gamma) \text{ and } v_z(f) \geq 0, \forall z \in \mathfrak{H}^*,$$

where the order $v_z(f)$ of f at a cusp $z = g(\infty)$ (with $g \in \Gamma(1)$) is defined via the Laurent expansion (2.4) of $f \circ g$ in terms of q_N, where $N = N_g$ (cf. Remark 2.3(d)).

(b) In order to be able to study the transformation properties of modular forms with respect to the action of $\mathrm{GL}_2^+(\mathbb{Q})$, it is useful to extend the notation $f|_k\alpha$ to matrices $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{Q})$ as follows:

$$f|_k\alpha = f \circ [\alpha]_k = (\det \alpha)^{k/2} f(\alpha(z))(cz + d)^{-k}.$$

Then for any $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ and $c \in \mathbb{Q}^+$, we have

$$f|_k(c\alpha) = f|_k\alpha. \tag{2.15}$$

(Note, however, that if $c < 0$ then we have $f \circ [c]_k = (-1)^k f$.) Moreover, the associative law (1.4) also holds for this extended symbol, i.e. for any $\alpha_1, \alpha_2 \in \mathrm{GL}_2^+(\mathbb{Q})$, we have

$$f|_k(\alpha_1\alpha_2) = (f|_k\alpha_1)|_k(\alpha_2). \tag{2.16}$$

The following properties of modular forms are easily verified (cf. [Ko], p. 127ff):

Proposition 2.20 *(a) $M_0(\Gamma) = \mathbb{C}$ and $M_k(\Gamma) = \{0\}$ if $k < 0$ or if k is odd and $-1 \in \Gamma$.*

(b) $M(\Gamma) := \oplus_{k\in\mathbb{Z}} M_k(\Gamma)$ is a graded ring and $S(\Gamma) := \oplus S_k(\Gamma)$ is a graded $M(\Gamma)$-ideal.

(c) If $f, g \in M_k(\Gamma)$ and $g \neq 0$, then $f/g \in \mathcal{M}(\Gamma)$.

(d) If $\Gamma_1 \leq \Gamma_2$ are subgroups, then $M_k(\Gamma_1) \supset M_k(\Gamma_2)$; in fact, $M_k(\Gamma_1)^{\Gamma_2} = M_k(\Gamma_2)$ and similarly, $S_k(\Gamma_1)^{\Gamma_2} = S_k(\Gamma_2)$.

(e) If $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, then the map $f \mapsto f|_k\alpha = f \circ [\alpha]_k$ defines isomorphisms

$$[\alpha]_k : M_k(\alpha\Gamma\alpha^{-1}) \xrightarrow{\sim} M_k(\Gamma), \quad [\alpha]_k : S_k(\alpha\Gamma\alpha^{-1}) \xrightarrow{\sim} S_k(\Gamma),$$

for any (generalized) congruence subgroup Γ. In particular, if $\alpha \in N_{\mathrm{GL}_2^+(\mathbb{Q})}(\Gamma)$ is in the normalizer of Γ in $\mathrm{GL}_2^+(\mathbb{Q})$, then $[\alpha]_k$ is an element of $\mathrm{Aut}_{\mathbb{C}}(M_k(\Gamma))$.

Corollary 2.21 *If Γ is a congruence subgroup and $k \in \mathbb{Z}$, then $\dim_{\mathbb{C}} M_k(\Gamma) < \infty$.*

Proof. This is trivial if $M_k(\Gamma) = \{0\}$, so assume that there is a non-zero modular form $f_0 \in M_k(\Gamma)$. Then by Proposition 2.20(c), $V := \frac{1}{f_0} M_k(\Gamma) \subset \mathcal{M}(\Gamma)$. Let $S = \{z \in \mathfrak{H}^* : v_z(f_0) > 0\}$ denote the set of zeros of f_0. Then S is Γ-stable and by an argument similar to that of the proof of Proposition 1.3 we see that $\Gamma\backslash S$ is a finite set. Now if $f \in V$, then $ff_0 \in M_k(\Gamma)$, so $v_z(ff_0) \geq 0, \forall z \in \mathfrak{H}^*$. Thus $v_z(f) \geq -v_z(f_0), \forall z \in S$ and $v_z(f) \geq 0, \forall z \notin S$, and so the following Lemma 2.22 shows that $\dim V < \infty$. Since $\dim M_k(\Gamma) = \dim V$, the assertion follows.

Lemma 2.22 *Let $S \subset \mathfrak{H}^*$ be a Γ-stable subset such that $\Gamma\backslash S$ is finite, and let $\nu : \Gamma\backslash S \to \mathbb{Z}$ be a function. Then the set*

$$L(S, \nu) := \{f \in \mathcal{M}(\Gamma) : v_z(f) \geq -\nu(z), \forall z \in S \text{ and } v_z(f) \geq 0, \forall z \in \mathfrak{H}^* \setminus S\}$$

is a finite-dimensional $\mathbb{C}$-vector space.

Proof. Clearly, $L(S, \nu)$ is a $\mathbb{C}$-vector space. Without loss of generality we may assume that $\nu(z) > 0, \forall z \in S$, for if $S' = \{z \in S : \nu(z) > 0\}$, then $L(S, \nu) \subset L(S', \nu_{|S'})$, and hence it is enough to verify the assertion for S' in place of S.

Let $z_1, \ldots, z_r$ be a system of representatives of $\Gamma\backslash S$, and put $n_i = \nu(z_i)$, $n = n_1 + \ldots + n_r$. At each z_i fix a local parameter t_i (i.e. $t_i = z - z_i$, if $z_i \in \mathfrak{H}$, and $t_i = q_{N_{z_i}}$, if $z_i \in \mathbb{Q} \cup \{\infty\}$), and write the Laurent expansion of $f \in \mathcal{M}(\Gamma)$ at z_i as $f = \sum_m a_{m,i}(f) t_i^m$. Consider the $\mathbb{C}$-linear map

$$T : L(S, \nu) \to \mathbb{C}^n$$

defined by the rule $f \mapsto (a_{-1,1}(f), \ldots, a_{-n_1,1}(f), a_{-1,2}(f), \ldots, a_{-n_r,r}(f)) \in \mathbb{C}^n$. Now if $f \in \mathrm{Ker}(T)$ then $v_{z_i}(f) \geq 0$, for $i = 1, \ldots, r$, and hence $v_z(f) \geq 0$, for all $z \in S$, since f is Γ-invariant. Moreover, since also $v_z(f) \geq 0$, for all $z \in \mathfrak{H}^* \setminus S$ by hypothesis, we see that $f \in M_0(\Gamma) = \mathbb{C}$, the latter by Proposition 2.20(a). Thus $\dim \mathrm{Ker}(T) \leq 1$, and so $\dim L(S, \nu) \leq n + 1$.

Example 2.23 (a) If $f \in \mathbf{M}_k = M_k(\Gamma(1))$, and $N \geq 1$ is an integer, then the form

$$(f \circ \beta_N)(z) \;=\; f(Nz) \;=\; N^{-k/2} f|_k \beta_N(z) \in M_k(\Gamma_0(N)),$$

for by Proposition 2.20(e), (b) we have $f \circ \beta_N \in M_k(\beta_N^{-1}\Gamma(1)\beta_N) \subset M_k(\Gamma(1) \cap \beta_N^{-1}\Gamma(1)\beta_N) = M_k(\Gamma_0(N))$, where the latter identity follows from (2.1). Similarly, if $f \in \mathbf{S}_k = S_k(\Gamma(1))$, then $f \circ \beta_N \in S_k(\Gamma_0(N))$. In particular, $\Delta_N = \Delta \circ \beta_N \in S_{12}(\Gamma_0(N)) \setminus S_{12}(\Gamma(1))$.

(b) *$\wp$-division values.* As in Example 2.4(c), let $\wp_{r,s,N}(\tau) = \wp(\frac{r\tau+s}{N}, \tau)$ be the N-division value of the Weierstrass $\wp$-function associated to the pair $(r,s) \in \mathbb{Z}^2$ with $(r,s) \not\equiv (0,0) \pmod N$. Then the discussion of Example 2.4(c) shows that $\wp_{r,s,N}$ is a modular form of weight 2 of level N, i.e. that $\wp_{r,s,N} \in M_2(\Gamma(N))$.

(c) *Eisenstein series.* Let $k \geq 3$ and fix an integer $N \geq 1$. For each $a = (a_1, a_2) \in \mathbb{Z}^2$ consider the *Eisenstein series*

$$G_k^a(z) = G_k^{a \bmod N}(z) = \sum_{\substack{m \,\in\, \mathbb{Z}^2 \\ m \,\equiv\, a \,(\mathrm{mod}\, N) \\ m \,\neq\, (0,0)}} \frac{1}{(m_1 z + m_2)^k}$$

which only depends on the image of a in $\mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$. It is immediate that this series converges absolutely on $\mathfrak{H}$ and hence defines a holomorphic function there. Furthermore we have:

0) If $a = (0,0)$, then $G_k^{a \bmod N}(z) = N^{-k} G_k(z) \in \mathbf{M}_k$.

1) For any $g \in \Gamma(1)$ we have $G_k^{a \bmod N}|_k g = G_k^{(ag) \bmod N}$.

2) Each $G_k^{a \bmod N}$ is holomorphic at ∞, i.e. G_k has an expansion in q_N with non-negative terms; cf. [Ko], p. 132 or [Sch], p. 156.

Thus, by the same argument as in Example 2.4(c) it follows from 1) and 2) that $G_k^{a \bmod N} \in M_k(\Gamma(N))$.

In fact, the G_k^a's are closely related to the division values of the derivatives $\wp^{(n)}(z,\tau) = \frac{d^n}{dz^n}\wp(z,\tau)$ of the Weierstrass $\wp$-function, for we have the formula:

$$G_k^{a \bmod N} \;=\; \frac{(-1)^k}{N^k (k-1)!} \wp^{(k-2)}\left(\frac{a_1\tau + a_2}{N}, \tau\right), \quad \forall \tau \in \mathfrak{H};$$

cf. [Ko], p. 134 or [Sch], p. 157.

(d) Although the Eisenstein series $E_2 = 1 - 24\sum_{n\geq 1}\sigma_1(n)q^n$ is *not* a modular form (of weight 2) for any congruence subgroup Γ (as the transformation law (1.11) shows), we can modify it slightly so that it becomes a modular form. More precisely, if $N > 1$ is any integer, then it follows from (1.11) that the function

$$E_{2,N}(z) = NE_2(Nz) - E_2(z) = (N-1) + 24\sum_{n\geq 1}\Big(\sum_{\substack{d \mid n \\ d \,\not\equiv\, 0 (N)}} d\Big) q^n$$

is a modular form of weight 2 on $\Gamma_0(N)$, i.e. $E_{2,N} \in M_2(\Gamma_0(N))$; cf. [Sch], p. 177.

Remark 2.24 (a) Fix integers N and $k \geq 3$ and let $E_k(\Gamma(N)) = \langle G_k^{a \bmod N} : a \in \mathbb{Z}^2 \rangle \subset M_k(\Gamma(N))$ denote the $\mathbb{C}$-vector space generated by the Eisenstein series. Then we have

$$M_k(\Gamma(N)) = E_k(\Gamma(N)) \oplus S_k(\Gamma(N)), \tag{2.17}$$

which follows easily from Theorem 2 of Schoeneberg[Sch], p. 158. (Note that if $N \leq 2$, then $(-1) \in \Gamma(N)$ and so $M_k(\Gamma(N)) = E_k(\Gamma(N)) = S_k(\Gamma(N)) = \{0\}$ when k is odd.) In addition, it follows from that theorem that

$$\dim E_k(\Gamma(N)) = \sigma_\infty(\Gamma(N)) := \#\mathrm{cusps}(\Gamma(N)) = \#(\Gamma(N)\backslash(\mathbb{Q} \cup \{\infty\})), \tag{2.18}$$

except in the case that $N \leq 2$ and $k \equiv 1\,(2)$ in which case $\dim E_k(\Gamma(N)) = 0$. Note that we have

$$\sigma_\infty(\Gamma(N)) = \frac{\mu(\Gamma(N))}{N}, \tag{2.19}$$

which follows easily from the fact that $\Gamma(1)/\pm\Gamma(N)$ acts transitively on the set of cusps of $\Gamma(N)$.

(b) Similarly, if we put $E_2(\Gamma(N)) = \langle \wp_{r,s,N} : (r,s) \in \mathbb{Z}^2, (r,s) \not\equiv (0,0) \pmod N \rangle$, then by Theorem 9 of Schoeneberg[Sch], p. 172, we see that the decomposition (2.17) also holds for $k = 2$. However, formula (2.18) is no longer true for $k = 2$; instead we have

$$\dim E_2(\Gamma(N)) = \sigma_\infty(\Gamma(N)) - 1.$$

(c) For any congruence subgroup Γ of level N, let us put $E_k(\Gamma) = E_k(\Gamma(N))^\Gamma$, if $k \geq 2$. It is then clear (by taking invariants of both sides of (2.17)) that the analogue of the decomposition (2.17) holds for Γ, i.e. that we have $M_k(\Gamma) = E_k(\Gamma) \oplus S_k(\Gamma)$.

(d) As in the case of $\Gamma = \Gamma(1)$, it turns out that the "Eisenstein space" $E_k(\Gamma)$ is in general only a small part of $M_k(\Gamma)$. For example, in the case of $\Gamma = \Gamma(N)$ we have the following general dimension formulae which hold for all $k \geq 3$ and $N \geq 3$:

$$\dim M_k(\Gamma(N)) = \mu\left(\frac{k-1}{12} + \frac{1}{2N}\right), \quad \dim S_k(\Gamma(N)) = \mu\left(\frac{k-1}{12} - \frac{1}{2N}\right), \tag{2.20}$$

where $\mu = \mu(\Gamma(N))$. (These and other dimension formulae will be discussed in more detail below.) Clearly, these spaces are much larger than $\dim E_k(\Gamma(N)) = \frac{\mu}{N}$. A similar statement is true for $k = 2$, except that in this case we have

$$\dim M_2(\Gamma(N)) = \mu\left(\frac{N+6}{12N}\right) \quad \text{and} \quad \dim S_2(\Gamma(N)) = \mu\left(\frac{N-6}{12N}\right) + 1. \tag{2.21}$$

Note that all these spaces grow quite rapidly with N because $\mu(\Gamma(N)) \approx N^3$; more precisely, we have the bounds

$$\frac{1}{2}N^3 > \mu(\Gamma(N)) > \frac{3}{\pi^2}N^3 = \frac{1}{2\zeta(2)}N^3.$$

Since these spaces grow so rapidly with N, it is useful to subdivide them further. In his works, Hecke made several suggestions to this end. Particularly useful is his decomposition of $M_k(\Gamma_1(N))$ into the subspaces $M_k(N,\chi)$ of *Nebentypus* χ which are defined as follows.

Proposition 2.25 *If $\Gamma = \Gamma_1(N)$, then we have the direct sum decompositions*

$$M_k(\Gamma_1(N)) = \oplus_\chi M_k(N,\chi) \quad \textit{and} \quad S_k(\Gamma_1(N)) = \oplus_\chi S_k(N,\chi),$$

in which the sums run over all characters $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ and

$$M_k(N,\chi) := \{f \in M_k(\Gamma_1(N)) :\ f|_k\gamma = \chi(d)f, \forall \gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)\},$$

and $S_k(N,\chi) := M_k(N,\chi) \cap S_k(\Gamma_1(N))$.

Proof. Recall from §2.2.1 that $\Gamma_1(N) \trianglelefteq \Gamma_0(N)$ and that the map $a \mapsto \sigma_a = \langle a \rangle \equiv \left(\begin{smallmatrix} a^{-1} & 0 \\ 0 & a \end{smallmatrix}\right) \pmod N$ induces an isomorphism $Z_N = (\mathbb{Z}/N\mathbb{Z})^\times \xrightarrow{\sim} \Gamma_0(N)/\Gamma_1(N)$. Thus the group Z_N acts on $M_k(\Gamma_1(N))$ (and on $S_k(\Gamma_1(N))$), and hence we have a decomposition of $M_k(\Gamma_1(N))$ into its χ-eigenspaces $M_k(\Gamma_1(N))_\chi = \{f \in M_k(\Gamma_1(N)) :\ f|_k\sigma_a = \chi(a)f\}$, where χ runs over all characters of Z_N. However, since $M_k(\Gamma_1(N))_\chi = M_k(N,\chi)$, we obtain the above decomposition of $M_k(\Gamma_1(N))$. The proof for $S_k(\Gamma_1(N))$ is similar.

Remark 2.26 (a) Since $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma_1(N)$, each $f \in M_k(\Gamma_1(N))$ has an expansion in $q = e^{2\pi iz}$, i.e.

$$f(z) = \sum_{n \geq 0} a_n(f) q^n, \quad \text{where } q = e^{2\pi iz}.$$

Conversely, if $f \in M_k(\Gamma(N))$ is a modular form of level N which has such an expansion, then $f \in M_k(\Gamma_1(N))$ because $\Gamma_1(N) = \langle \Gamma(N), T \rangle$.

(b) Since $-1 \in \Gamma_0(N)$, we see from the definition that $M_k(N,\chi) = 0$ if $\chi(-1) \neq (-1)^k$.

(c) If $\chi = 1$ is the trivial character, then $M_k(N,1) = M_k(\Gamma_0(N))$ by definition. Hecke called these forms the *Haupttypus* (main type) and the modular forms with $\chi \neq 1$ forms of *Nebentypus* (auxiliary type) χ.

We now consider some examples of modular forms of type (N,χ) which occur naturally in number theory.

Example 2.27 (a) *Theta-series.* As in §1.2.5, let $Q(\vec{x}) = \frac{1}{2}\vec{x}^t A \vec{x}$ be an even, integral, positive definite quadratic form in $r = 2k$ variables ($k \in \mathbb{Z}$), and let

$$\vartheta_Q(z) = \sum_{\vec{m} \in \mathbb{Z}^r} q^{Q(\vec{m})} = \sum_{\vec{m} \in \mathbb{Z}^r} e^{\pi i z \vec{m}^t A \vec{m}}$$

be the associated theta-series. Let N be the smallest integer $M > 0$ such that MA^{-1} is an even integral matrix; in other words,

$$N = \frac{D}{g} \quad \text{where } D = |\det(A)| \quad \text{and} \quad g = \gcd\left(\{a_{ij}, \tfrac{1}{2}a_{ii}\}_{1 \leq i \leq j \leq r}\right);$$

cf. [Sch], p. 207. Moreover, define the quadratic character $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \{\pm 1\}$ by

$$\chi(d) = \mathrm{sign}(d)^k \left(\frac{(-1)^k D}{|d|}\right),$$

where $\left(\frac{\cdot}{\cdot}\right)$ denotes the Jacobi-Kronecker symbol; cf. [Hua], p. 304. Then it turns out (cf. [Sch], p. 217-218 or [Iw], p. 175) that

$$\vartheta_Q \in M_k(N, \chi), \quad \text{where } N \text{ and } \chi \text{ are as above.} \tag{2.22}$$

For example, if $Q(x_1, x_2) = ax_1^2 + bx_1x_2 + cx_2^2$ is a primitive positive definite binary quadratic form, i.e. if $a > 0$, $b^2 - 4ac < 0$, $a, b, c \in \mathbb{Z}$ and $\gcd(a, b, c) = 1$, then $N = D = 4ac - b^2$ and hence $\vartheta_Q \in M_1(N, \chi)$, where $\chi = \chi_{-D} = \left(\frac{-D}{\cdot}\right)$.

It is perhaps interesting to sketch some of the ideas involved in the proof of (2.22). For this it seems necessary to consider more generally the *congruent theta-series*

$$\vartheta_{Q,\vec{x}}(z) \ := \sum_{\vec{m} \in \mathbb{Z}^r} q^{Q(\vec{m} + \frac{\vec{x}}{N})} \ = \sum_{\substack{\vec{m} \in \mathbb{Z}^r \\ \vec{m} \equiv \vec{x} \, (mod\, N)}} e^{\pi i z (\vec{m}^t A \vec{m})/N^2}, \quad \text{where } \vec{x} \in \mathbb{Z}^r.$$

Then by using the Poisson summation formula one obtains the inversion formula

$$\vartheta_{Q,\vec{x}}(\tau) = \frac{1}{\sqrt{D}} \left(\frac{i}{\tau}\right)^k \sum_{\vec{m} \in \mathbb{Z}^r} \mathbf{e}\left(\frac{-\vec{m} A^{-1} \vec{m}}{2\tau} + \vec{m}^t \vec{x}\right)$$

in which $\mathbf{e}(z) = e^{2\pi i z}$; cf. [Iw], p. 167. (This generalizes the transformation formula (1.28) of §1.2.5.)

We now restrict attention to vectors $\vec{x} \in G(Q) := \{\vec{x} \pmod N : A\vec{x} \equiv 0 \pmod N\}$, which is a finite group of order D; cf. [Iw], p. 168. (Note that since $\vartheta_{Q,\vec{x}} = \vartheta_{Q,\vec{x}}$ if $\vec{x} \equiv \vec{y} \pmod N$, the function $\vartheta_{Q,\vec{x}}$ is well-defined for any $\vec{x} \in \mathbb{Z}^r/N\mathbb{Z}^r$.) For such an $\vec{x}$ it follows from the inversion formula that we have:

$$\vartheta_{Q,\vec{x}}|_k T = \psi_Q(\vec{x}) \vartheta_{Q,\vec{x}} \quad \text{and} \quad \vartheta_{Q,\vec{x}}|_k S = \frac{(-i)^k}{\sqrt{D}} \sum_{\vec{y} \in G(Q)} \psi_Q(\vec{x}, \vec{y}) \vartheta_{Q,\vec{y}} \tag{2.23}$$

in which $\psi_Q(\vec{x}) = e^{2\pi i Q(\vec{x}/N)}$ and $\psi_Q(\vec{x}, \vec{y}) = \psi_Q(\vec{x} + \vec{y}) \psi_Q(\vec{x})^{-1} \psi_Q(\vec{y})^{-1} = e^{2\pi i (\vec{x}^t A \vec{y})/N^2}$; cf. [Iw], p. 169-170 or [Sch], p. 210. From this we see that the $\mathbb{C}$-vector space $V_Q := \langle \vartheta_{Q,\vec{x}} : \vec{x} \in G(Q) \rangle$ is stable under the action of $\Gamma(1)$. In particular, since each $\vartheta_{Q,\vec{x}}$ has a power series expansion in q_N, it follows that each $\vartheta_{Q,\vec{x}}$ is holomorphic at all the cusps.

With more work it is possible to deduce from the above transformation laws (2.23) the rule

$$\vartheta_{Q,\vec{x}}|_k g \ = \ \chi(d) \psi_Q(\vec{x})^{ab} \vartheta_{Q,a\vec{x}}, \quad \text{if } \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N); \tag{2.24}$$

cf. [Sch], p. 218. Since $\psi_Q(\vec{x})$ is an N-th root of unity, we see from (2.24) that $\vartheta_{Q,\vec{x}} \in M_k(\Gamma(N)), \forall \vec{x} \in G(Q)$, and that $\vartheta_Q = \vartheta_{Q,\vec{0}} \in M_k(N, \chi)$.

(b) *Dirichlet L-functions.* Let χ be a *primitive Dirichlet character* mod N, i.e. $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ is a homomorphism with $\chi(1 + M\mathbb{Z}/N\mathbb{Z}) \neq \{1\}$ for any proper divisor $M|N$. If we lift χ to a multiplicative map $\chi : \mathbb{Z} \to \mathbb{C}$ by setting $\chi(a) = 0$ if $(a, N) > 1$, then the associated Dirichlet L-function is defined by

$$L(s, \chi) = \sum_{n \geq 1} \frac{\chi(n)}{n^s} = \prod_{p \nmid N} \left(1 - \chi(p)p^{-s}\right)^{-1}.$$

For example, if $\chi = 1$ is the trivial character, then $N = 1$ and $L(s, \chi) = \zeta(s)$ is just the Riemann zeta-function.

Now let χ_1 and χ_2 be two primitive Dirichlet characters mod N_1 and N_2, respectively, and put $N = N_1N_2$ and $\chi = \chi_1\chi_2$. Fix an integer $k \geq 1$ satisfying $\chi(-1) = (-1)^k$, and put

$$a_n = a_n(\chi_1, \chi_2, k) = \sum_{d|N} \chi_1(n/d)\chi_2(d)d^{k-1} \quad \text{for } n \geq 1.$$

Then there is a unique constant a_0 (which is given explicitly on p. 177 of [Mi]) such that

$$f = f_{\chi_1,\chi_2,k} := \sum_{n=0}^{\infty} a_n q^n \in M_k(N, \chi);$$

cf. [Mi], p. 177. Thus, f is the unique modular form such that its associated L-function $L(f, s)$ (cf. §1.5.2) is given by the formula

$$L(f, s) = L(s, \chi_1)L(s - k + 1, \chi_2).$$

For example, in the case that $\chi_1 = \chi_2 = 1$ (and hence $N_1 = N_2 = N = 1$), $f(z) = E_k(z)$ as we already saw in Example 1.34. In fact, it turns out that any such f is always a generalized Eisenstein series, i.e. $f \in E_k(\Gamma_1(N))$; cf. [Mi], p. 179.

(c) *Hecke L-functions.* In 1918 Hecke gave a vast generalization of Dirichlet characters and Dirichlet L-functions to arbitrary number fields K by introducing *Grössencharacters* (often called *Hecke characters*); cf. [Mi], p. 91. In the case that $K = \mathbb{Q}(\sqrt{-d})$ is an imaginary quadratic field, such Grössencharacters are defined as follows.

Fix an integer $r \geq 0$ and an ideal $\mathfrak{m}$ of the ring $\mathfrak{O}_K$ of integers of K, and let $I(\mathfrak{m})$ denote the group of fractional ideals of $\mathfrak{O}_K$ which are prime to $\mathfrak{m}$. A homomorphism

$$\psi : I(\mathfrak{m}) \to \mathbb{C}_1 := \{z \in \mathbb{C} : |z| = 1\}$$

is called a *Grössencharacter* mod $\mathfrak{m}$ *of type* r if ψ satisfies the condition:

$$\psi((a)) = \left(\frac{a}{|a|}\right)^r, \quad \forall a \equiv 1 \pmod{\mathfrak{m}}.$$

The associated *Hecke L-function* is defined by

$$L(s, \psi) = \sum_{\substack{\mathfrak{a} \\ (\mathfrak{a}, \mathfrak{m}) = 1}} \frac{\psi(\mathfrak{a})}{(N\mathfrak{a})^s} = \prod_{\mathfrak{p} \nmid \mathfrak{m}} \left(1 - \psi(\mathfrak{p})(N\mathfrak{p})^{-s}\right)^{-1}$$

in which $N\mathfrak{a} = \#(\mathfrak{O}_K/\mathfrak{a})$ denotes the norm of an $\mathfrak{O}_K$-ideal $\mathfrak{a}$. Clearly, $L(s,\psi)$ is a Dirichlet series:

$$L(s,\psi) = \sum_{n\geq 1} \frac{a_n(\psi)}{n^s} \quad \text{with } a_n(\psi) = \sum_{N\mathfrak{a}=n} \psi(\mathfrak{a}).$$

For example, if $\psi = 1$ (and $\mathfrak{m} = (1)$, $r = 0$), then $L(s,\psi) = \zeta_K(s)$ is just the Dedekind ζ-function of K. Note also that a Grössencharacter mod $\mathfrak{m} = (1)$ of type $r = 0$ is the same as a character on the ideal class group $Cl(\mathfrak{O}_K)$.

Hecke showed that such L-functions come from modular forms. More precisely, if we put

$$f_\psi = \sum_{n\geq 1} a_n(\chi) q^n,$$

then by [Mi], p. 183, we have that

$$f_\psi \in M_{r+1}(N,\chi), \quad \text{where } N = |d_K|(N\mathfrak{m})$$

and d_K is the discriminant of K and where χ is the Dirichlet character mod N defined by the formula

$$\chi(n) = \chi_K(n)\psi((n)), \quad \text{if } n \in \mathbb{Z},\ (n,N) = 1$$

in which $\chi_K = \left(\frac{d_K}{\cdot}\right)$ denotes the quadratic character associated to K (given by the Jacobi-Kronecker symbol). In fact, f_ψ is a always a cusp form (i.e. $f_\chi \in S_{r+1}(N,\chi)$) except when $r = 0$ and $\psi = \chi' \circ N_{K/\mathbb{Q}}$ for some Dirichlet character χ'; cf. [Mi], p. 183.

(d) *Twists of cusp forms.* Let $f \in S_k(N,\chi)$ be a cusp form of weight k of type (N,χ) and let ψ be a (primitive) Dirichlet character mod M. Put

$$f_\psi = \sum_{n=1}^{\infty} \psi(n) a_n(f) q^n, \quad \text{where } f = \sum_{n=1}^{\infty} a_n(f) q^n \text{ is the } q\text{-expansion of } f.$$

Then f_ψ is again a cusp form, but of a different type. More precisely, we have that

$$f_\psi \in S_k(\tilde{N}, \chi\psi^2), \quad \text{where } \tilde{N} = \mathrm{lcm}(N, M^2, NM). \tag{2.25}$$

To see this, observe first that we have the identity

$$\sum_{a=1}^{M} \overline{\psi}(a) f|_k \xi_{a,M} \ = \ g(\overline{\psi}) f_\psi \quad \text{in which } g(\overline{\psi}) = \sum_{a=1}^{M} \overline{\psi}(a) e^{2\pi i a/M}$$

denotes the *Gauss sum* associated to $\overline{\psi}$ and $\xi_{a,M} = \left(\begin{smallmatrix} 1 & \frac{a}{M} \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Q})$. Since the left hand side of this identity is easily seen to be in $S_k(\tilde{N}, \chi\psi^2)$ (cf. [Sh], p. 92) and since $g(\overline{\psi}) \neq 0$ (cf. [Sh], p. 91), if follows that also $f_\psi \in S_k(\tilde{N}, \chi\psi^2)$.

Note that it follows from (2.25) that if $f \in S_k(\Gamma_1(N))$ is any cusp form and $M > 1$ is any integer, then its "prime to M projection"

$$f_{(M)} = \sum_{\substack{n\geq 1 \\ (n,M)=1}} a_n(f) q^n$$

is also a cusp form of some level. Indeed, choose a primitive Dirichlet character ψ mod M', where $M' = M$, if $M \neq 2$, and $M' = 4$ if $M = 2$. Then by the above we have

$$f_{(M)} = (f_\psi)_{\overline{\psi}} \in S_k(\Gamma_1(\tilde{N})) \quad \text{with } \tilde{N} = \mathrm{lcm}(N, (M')^2, NM').$$

(e) *L-functions of elliptic curves.* Let $E/\mathbb{Q}$ be an elliptic curve, i.e. E is defined by an equation of the form

$$y^2 = x^3 + ax + b \quad \text{with } a, b \in \mathbb{Q} \text{ and } 4a^3 + 27b^2 \neq 0.$$

By replacing E by an isomorphic curve (i.e. by replacing a by ac^4 and b by bc^6 with $c \in \mathbb{Q}$) we can assume that $a, b \in \mathbb{Z}$ and that the absolute value of the discriminant $\Delta_E = -16(4a^3 + 27b^2) \in \mathbb{Z}$ is minimal (among all such choices). For each prime $p \nmid \Delta_E$ put

$$N_p(E) = \#\{(x, y) \bmod p : y^2 \equiv x^3 + ax + b\} \quad \text{and} \quad a_p(E) = p - N_p(E),$$

and consider the function

$$L_E^*(s) = \prod_{p \nmid \Delta_E} (1 - a_p(E)p^{-s} + p^{1-2s})^{-1}$$

which is called the *Hasse-Weil L-function* of $E/\mathbb{Q}$. Note that this product converges for $\mathrm{Re}(s) > \frac{3}{2}$ because by a theorem of Hasse we have $|a_p(E)| < 2\sqrt{p}$.

Hasse conjectured that $L_E^*(s)$ has a functional equation and an analytic continuation to all of $\mathbb{C}$, and this was verified by Weil [We1]) in 1952 in a few cases. For these cases, Weil was able to identify $L_E^*(s)$ with a Hecke L-function associated to the Grössencharacter defined by certain Jacobi sums. This was then generalized by Deuring in 1953 to all elliptic curves $E/\mathbb{Q}$ *with complex multiplication.* The latter are elliptic curves which have an analytic description of the form $E \simeq \mathbb{C}/\mathfrak{O}_K$, where K is an imaginary quadratic number field (such that $\mathfrak{O}_K$ has unique factorization). For example, the curves of the form

$$y^2 = x^3 + ax \quad \text{and} \quad y^2 = x^3 + b$$

are such curves of complex multiplication. (The first family (with $b = 0$) is analytically isomorphic to $\mathbb{C}/\mathbb{Z}[i]$ and the second (with $a = 0$) is isomorphic to $\mathbb{C}/\mathbb{Z}[e^{2\pi i/3}]$.) For these curves, the associated Grössencharacters and L-functions are described in Ireland/Rosen[IR], chapter 18. (See also [Ko], chapter 2). The general case is treated in Silverman[Si2], chapter 2.

In 1955 Taniyama formulated the following remarkable conjecture:

Conjecture (Taniyama). *For every elliptic curve $E/\mathbb{Q}$, its Hasse-Weil L-function $L_E^*(s)$ comes from a modular form of weight 2 on some congruence subgroup Γ.*

More precisely, he stated his conjecture as follows: "If Hasse's Conjecture is correct, then L_E^* has to come from an automorphic form." This statement was made more precise by Weil[We2] who proved in 1967 the following generalization of Hecke's result:

If $f = \sum a_n(f)q^n$ is a holomorphic function on $\mathfrak{H}$ such that $L(f,s)$ is bounded in vertical strips and such that $L(f_\chi, s)$ has a functional equation (of the right type) for every primitive Dirichlet character χ, then f is a cusp form (of some level).

Actually, this result cannot be applied directly to $L_E^*(s)$ since it doesn't satisfy the right functional equation. However, by introducing certain Euler factors for the primes $p|\Delta_E$ as well, Weil defined the (refined) Hasse-Weil L-function $L_E(s)$ which satisfies (at least in special cases) the right functional equation.

In his book, Shimura[Sh] gave a general construction of *all* the examples which satisfy Taniyama's Conjecture. More precisely, if we start with $f \in S_2(\Gamma_0(N))$ such that its L-function has an Euler product (of the right type), then Shimura's construction finds an elliptic curve $E/\mathbb{Q}$ such that $L_E(s) = L(f,s)$. In particular, $L_E(s)$ satisfies Hasse's Conjecture.

In 1995 Wiles[Wi] proved a substantial part of Taniyama's Conjecture, and his method was refined by Breuil, Conrad, Diamond and Taylor[BCDT] to yield a complete proof of Taniyama's Conjecture:

$$\text{For every } E/\mathbb{Q} \text{ there exists } f_E \in S_2(\Gamma_0(|\Delta_E|)) \quad \text{such that} \quad L(f_E, s) = L_E(s).$$

Actually, it turns out that f_E has much smaller level than $|\Delta_E|$, for the proof shows that $f_E \in S_2(\Gamma_0(N_E))$, where $N_E|\Delta_E$ is the conductor of the elliptic curve (in which the prime divisors $p \neq 2, 3$ of Δ_E appear with multiplicity at most 2).

A very important property of the spaces $M_k(\Gamma)$ and $S_k(\Gamma)$ is that their dimension can be computed explicitly in terms of basic group-theoretical data. To this end we introduce the following notation.

Notation. If Γ is a congruence subgroup, and $n > 1$ is an integer, then we put

$$\varepsilon_n(\Gamma) = \#\{z \in \Gamma\backslash\mathfrak{H} : \#(\mathrm{Stab}_\pm(z)/(\pm 1)) = n\},$$

where (as usual) $\mathrm{Stab}_{\pm\Gamma}(z) := \{g \in \Gamma : g(z) = z\}$ denotes the *stabilizer* of $z \in \mathfrak{H}$ with respect to $\pm\Gamma$. Thus, since a point $z \in \mathfrak{H}$ is called an *elliptic point* on $\mathfrak{H}$ with respect to Γ if $\mathrm{Stab}_{\pm\Gamma}(z) \neq \{\pm 1\}$, the number $\varepsilon_n(\Gamma)$ represents the number of Γ-inequivalent elliptic points on $\mathfrak{H}$ of order n.

Proposition 2.28 *If Γ is a congruence subgroup, then the dimension of $S_2(\Gamma)$ is given by the formula*

$$\dim S_2(\Gamma) = g_\Gamma = 1 + \frac{\mu(\Gamma)}{12} - \frac{\varepsilon_2(\Gamma)}{4} - \frac{\varepsilon_3(\Gamma)}{3} - \frac{\sigma_\infty(\Gamma)}{2}. \tag{2.26}$$

where, as before, $\mu(\Gamma) = [\Gamma(1) : \pm\Gamma]$ and $\sigma_\infty(\Gamma) = \#\mathrm{cusps}(\Gamma)$.

Proof (Sketch). The map $f \mapsto \omega_f = f(z)dz$ defines an isomorphism

$$\omega_\Gamma : S_2(\Gamma) \xrightarrow{\sim} \Omega^1(X_\Gamma)$$

from the space of weight 2 cusp forms on Γ to the space of holomorphic differential forms on the compact Riemann surface X_Γ; cf. [Sh], p. 39. Now by the *Riemann-Roch Theorem* we have $\dim_{\mathbb{C}} \Omega^1(X_\Gamma) = g_\Gamma$, where g_Γ denotes the *genus* of the compact Riemann surface X_Γ; cf. [Sh], p. 36. By using the *Riemann-Hurwitz* formula, one finds that the genus of X_Γ is given by formula of (2.26); cf. [Sh], p. 23 or [Mi], p. 113.

Remark 2.29 (a) For any $k \geq 2$, one can express $\dim M_k(\Gamma)$ and $\dim S_k(\Gamma)$ in terms of the genus of X_Γ and the invariants $\varepsilon_2, \varepsilon_3$ and σ_∞; cf. [Sh], p. 46, or [Mi], p. 60. However, no formula is known for $k = 1$.

(b) For $\Gamma = \Gamma(N)$ and $N \geq 2$ there are no elliptic points on Γ, i.e. $\varepsilon_2(\Gamma(N)) = \varepsilon_3(\Gamma(N)) = 0$. Thus, by (2.26) and (2.19) the formula for the genus of $X_{\Gamma(N)}$ becomes

$$g_{\Gamma(N)} = \mu(\Gamma(N)) \left(\frac{N-6}{12N} \right) + 1.$$

From this, together the results in [Sh], pp. 46-47, the explicit formulae of Remark 2.24(d) follow immediately.

(c) Similarly, the group $\Gamma = \Gamma_1(N)$ has no elliptic elements for $N \geq 4$, but the number of cusps is given by a more complicated formula; cf. [Mi], p. 111. If $N = p \geq 5$ is a prime then $\sigma_\infty = p - 1$ and $\mu = \frac{1}{2}(p^2 - 1)$, and so we obtain

$$g_{\Gamma_1(p)} = \tfrac{1}{24}(p-1)(p-11) + 1.$$

(d) On the other hand, the group $\Gamma = \Gamma_0(N)$ usually does have elliptic elements. The numbers $\varepsilon_2(\Gamma), \varepsilon_3(\Gamma)$ and $\sigma_\infty(\Gamma)$ are known explicitly (cf. [Mi], p. 108), but they more complicated than those of $\Gamma(N)$ since they depend on the Legendre symbol of the primes dividing N.

For example, if $N = p$ is an odd prime, then $\varepsilon_2 = 1 + \left(\frac{-1}{p}\right)$, $\varepsilon_3 = 1 + \left(\frac{-3}{p}\right)$ and $\sigma_\infty = 2$ (cf. [Mi], p. 108) and so (since $\mu = p + 1$) we obtain

$$g_{\Gamma_0(p)} = \begin{cases} [\frac{p+1}{12}] & \text{if } p \not\equiv 1 \pmod{12} \\ [\frac{p+1}{12}] - 1 & \text{if } p \equiv 1 \pmod{12} \end{cases}.$$

In particular, we see that $g_{\Gamma_0(p)} \sim \frac{p}{12}$ as $p \to \infty$.

Example 2.30 (a) For $\Gamma = \Gamma_0(11)$ we see from the genus formula of Remark 2.29(c) that $g_\Gamma = 1$, and hence $\dim_{\mathbb{C}} S_2(\Gamma_0(11)) = 1$ by Proposition 2.28. Now the function $g(z) = \eta(z)\eta(11z) \in S_2(\Gamma_0(11))$ (cf. [Ko], p. 130), and so $S_2(\Gamma_0(11)) = \mathbb{C}\eta(z)\eta(11z)$.

More generally, for any $N|12$, $N > 1$ we have $S_k(\Gamma_0(N-1)) = \mathbb{C}(\eta(z)\eta((N-1)z))^k$, if $k = \frac{24}{N}$; cf. [Sh], p. 49.

(b) For any $N|12$, we have $S_k(\Gamma(N)) = \mathbb{C}\eta(z)^{2k}$, if $k = \frac{12}{N}$; cf. [Sh], p. 50. In particular, $S_1(\Gamma(12)) = \mathbb{C}\eta(z)^2$, so $\eta(z)^2$ is a cusp form of weight 1 (and of level 12).

2.3 Hecke Operators

As we saw in Chapter 1, the Hecke operators T_n and the resulting Hecke algebra $\mathbb{T}$ played a fundamental role in the theory of modular forms of level 1. The same is true of higher level, except that the theory is somewhat more complicated.

The Hecke operators T_n are special cases of a slightly more general class of operators $T(\alpha)$. To define these, let Γ_1 and Γ_2 be two congruence subgroups and let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. We then define the linear map

$$T_{\Gamma_1,\Gamma_2}(\alpha) : M_k(\Gamma_1) \to M_k(\Gamma_2)$$

by the rule

$$f|_k T_{\Gamma_1,\Gamma_2}(\alpha) = (\det \alpha)^{k/2-1} \mathrm{tr}_{\Gamma_\alpha/\Gamma_2}(f|_k\alpha),$$

where $\Gamma_\alpha = \Gamma_2 \cap \alpha^{-1}\Gamma_1\alpha$ and where the *trace map* $tr = tr_{\Gamma_\alpha/\Gamma_2} : M_k(\Gamma_\alpha) \to M_k(\Gamma_2)$ is defined by the formula

$$\mathrm{tr}_{\Gamma_\alpha/\Gamma_2}(f) = \sum_{\gamma_i \in \Gamma_\alpha\backslash\Gamma_2} f|_k\gamma_i \in M_k(\Gamma_2), \quad \text{for } f \in M_k(\Gamma_\alpha). \tag{2.27}$$

(Note that the right hand side of (2.27) is independent of the choice of the coset representatives $\{\gamma_i\}$ of $\Gamma_\alpha\backslash\Gamma_2$.) In other words, the map $T(\alpha) = T_{\Gamma_1,\Gamma 2}(\alpha)$ is defined by means of the following commutative diagram (in which $d = \det(\alpha)^{k/2-1}$):

$$\begin{array}{ccc} M_k(\alpha\Gamma_\alpha\alpha^{-1}) & \xrightarrow{[\alpha]_k} & M_k(\Gamma_\alpha) \\ \uparrow & & \downarrow d\cdot tr \\ M_k(\Gamma_1) & \xrightarrow{T(\alpha)} & M_k(\Gamma_2) \end{array}$$

Remark 2.31 (a) The operator $T_{\Gamma_1,\Gamma_2}(\alpha)$ only depends on the double coset $\Gamma_1\alpha\Gamma_2$ defined by α. More precisely, if $\Gamma_1\alpha\Gamma_2 = \dot\cup\Gamma_1\alpha_i$ is any decomposition of $\Gamma_1\alpha\Gamma_2$ into Γ_1-cosets, then we have the formula

$$f_1|_k T_{\Gamma_1,\Gamma_2}(\alpha) = (\det \alpha)^{k/2-1} \sum f_1|_k\alpha_i, \quad \text{for all } f_1 \in M_k(\Gamma_1). \tag{2.28}$$

[Indeed, we see easily that the right hand side of (2.28) does not depend on the choice of coset representatives $\{\alpha_i\}$, and hence (2.28) follows because $\Gamma_1\alpha\Gamma_2 = \dot\cup\Gamma_1\alpha\gamma_i$ if $\Gamma_2 = \dot\cup\Gamma_\alpha\gamma_i$; cf. [Sh], p. 51.] Thus, $T_{\Gamma_1,\Gamma_2}(\alpha)$ coincides with the operator $[\Gamma_1\alpha\Gamma_2]_k$ defined by [Sh], p. 73. (Note that Koblitz[Ko] (p. 166) does not include the factor $\det(\alpha)^{k/2-1}$ in his definition of $[\Gamma_1\alpha\Gamma_2]_k$.)

(b) The following properties of $T(\alpha) = T_{\Gamma_1,\Gamma_2}(\alpha)$ are easily verified; cf. [Sh], p. 73ff or [Ko], p. 165ff:

1) If $c \in \mathbb{Q}$ and $c > 0$, then $T(c\alpha) = c^{k-2}T(\alpha)$.

2) If $f \in S_k(\Gamma_1)$, then $f|_k T_{\Gamma_1,\Gamma_2}(\alpha) \in S_k(\Gamma_2)$.

3) If $\alpha \in N_{SL_2(\mathbb{Q})}(\Gamma)$, then $f|_k T_{\Gamma,\Gamma}(\alpha) = f|_k\alpha$.

3′) If $\alpha \in N_{SL_2(\mathbb{Q})}(\Gamma)$, then for all $\beta \in \mathrm{GL}_2^+(\mathbb{Q})$ we have $f|_k T_{\Gamma,\Gamma}(\alpha\beta) = (f|_k\alpha)|_k T_{\Gamma,\Gamma}(\beta)$ and $f|_k T_{\Gamma,\Gamma}(\beta\alpha) = (f|_k T_{\Gamma,\Gamma}(\beta))|_k\alpha$.

4) The composition $T_{\Gamma_2,\Gamma_3}(\beta) \circ T_{\Gamma_1,\Gamma_2}(\alpha)$ of two such operators is computed as follows. Write

$$(\Gamma_1\alpha\Gamma_2)(\Gamma_2\beta\Gamma_3) = \dot{\bigcup} \Gamma_1\gamma_i\Gamma_3,$$

and let $n_i = \#\{\Gamma_2\delta_i : \Gamma_2\delta_i \subset \Gamma_2\beta\Gamma_3 \cap \Gamma_2\alpha^{-1}\Gamma_1\gamma_i\}$ denote the number of left cosets in $\Gamma_2\beta\Gamma_3 \cap \Gamma_2\alpha^{-1}\Gamma_1\gamma_i$. Then we have (cf. [Sh], p. 74 and pp. 51-52):

$$(f|_k T_{\Gamma_1,\Gamma_2}(\alpha))|_k T_{\Gamma_2,\Gamma_3}(\beta) = \sum_i n_i f|_k T_{\Gamma_1,\Gamma_3}(\gamma_i)$$

We now specialize to the case $\Gamma_1 = \Gamma_2 = \Gamma_1(N)$ and write $\Gamma = \Gamma_1(N)$. If $n = p$ is a prime, then the *Hecke operator* T_p (or $T(p)$) is defined by

$$T_p = T(p) = T_{\Gamma,\Gamma}(\alpha_p), \quad \text{where (as before) } \alpha_p = \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right).$$

More generally, for an arbitrary positive integer n define as in [Sh], p. 70,

$$T_{1,n} = T(1,n) = T_{\Gamma,\Gamma}(\alpha_n) \quad \text{and} \quad T_n = T(n) = \sum_{\Gamma\alpha\Gamma \subset \Delta'_n} T_{\Gamma,\Gamma}(\alpha),$$

where $\Delta'_n = \{\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in M_2(\mathbb{Z}) : \det\alpha = n,\ a \equiv 1 \pmod N, c \equiv 0 \pmod N\}$. Note that this definition agrees with the previous one when $n = p$ is a prime because $\Delta'_p = \Gamma\alpha_p\Gamma$; cf. Proposition 2.32(f) below.

In addition to the T_n's, it is also useful define the operators $T_{n,n} = T(n,n)$ by

$$T_{n,n} = T(n,n) = T_{\Gamma,\Gamma}(n\sigma_n), \quad \text{if } \gcd(n,N) = 1;$$

cf. [Sh], p. 72. Here $\sigma_n \in \Gamma_0(N)$ is as in the proof of Proposition 2.25, i.e.

$$\sigma_n \equiv \begin{pmatrix} n^{-1} & 0 \\ 0 & n \end{pmatrix} \pmod N.$$

The Hecke operators T_n satisfy the following fundamental properties:

Proposition 2.32 *(a) If m and n are coprime, then $T_{nm} = T_n \circ T_m$.*

(b) If $p|N$, then $T_{p^r} = (T_p)^r$.

(c) If $p \nmid N$ and $r \geq 2$, then $T_{p^r} = T_{p^{r-1}}T_p - pT_{p^{r-2}}T_{p,p}$.

(d) If $\gcd(a,N) = 1$, then the operators T_n and $T_{a,a}$ commute.

(e) For any positive integers m and n the operators T_n and T_m commute, and we have

$$T_n \circ T_m = \sum_{\substack{d|(m,n)\\(d,N)=1}} dT_{d,d}T_{mn/d^2}.$$

(f) If n is squarefree, then $T_n = T_{1,n}$. More generally, for any $n \geq 1$ we have

$$T_n = \sum_{\substack{d^2|n\\(d,\,N)\,=\,1}} T_{d,d}T_{1,n/d^2}.$$

Proof. (a) – (d): [Ko], p. 156; (e): [Sh], equation (3.3.6), p. 71.

(f) Let $\Delta_n^* = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Delta_n' : \gcd(a,b,c,d) = 1\}$. Then one easily sees (cf. [Ko], Lemma, p. 167) that $\Delta_n^* = \Gamma\alpha_n\Gamma$, and so we obtain the double coset decomposition

$$\Delta_n' = \bigcup_{\substack{d^2|n\\(d,\,N)\,=\,1}} d\sigma_d\Delta_{n/d^2}^* = \bigcup_{\substack{d^2|n\\(d,\,N)\,=\,1}} \Gamma d\sigma_d\alpha_{n/d^2}\Gamma,$$

from which the assertion follows.

Remark 2.33 (a) The above properties (b) – (d) may be summarized by following formal identity:

$$\sum_{n=1}^{\infty} T(n)n^{-s} = \prod_{p|N}(1 - T_p p^{-s})^{-1} \prod_{p\nmid N}(1 - T_p p^{-s} + T_{p,p}p^{1-2s})^{-1}.$$

(b) Since $\sigma_n \in N_{SL_2(\mathbb{Q})}(\Gamma_1(N))$, it follows from the definition and the properties of Remark 2.31(b) that

$$f|_k T_{n,n} = n^{k-2} f|_k \sigma_n, \quad \text{for all } f \in M_k(\Gamma_1(N)); \tag{2.29}$$

in particular,

$$f|_k T_{n,n} = n^{k-2}\chi(n)f, \quad \text{for } f \in M_k(N,\chi).$$

Thus, since Hecke operator T_m commutes with $T_{n,n}$ (cf. Proposition 2.29(d)), it follows that T_m maps $M_k(N,\chi)$ into itself. Indeed, if $f \in M_k(N,\chi)$, then

$$(f|_k T_m)|\sigma_n = n^{2-k}(f|_k T_m)|_k T_{n,n} = n^{2-k}(f|_k T_{n,n})|_k T_m = \chi(n)f|_k T_m, \quad \forall (n,N) = 1,$$

and so $f|_k T_m \in M_k(N,\chi)$.

We now examine the effect of the Hecke operators on the q-expansion of modular forms $f \in M_k(\Gamma_1(N))$. For this, we first introduce the following operators U_n and V_n on formal power series:

Notation. If $f = \sum a_n q^n \in \mathbb{C}[[q]]$, then put

$$V_m f = \sum_m a_n q^{mn} \quad \text{and} \quad U_m f = a_n q^{n/m},$$

where the second summation is only over those n which are divisible by m. Thus

$$U_1 f = V_1 f = U_m V_m f = f, \quad \text{whereas} \quad V_m U_m f = \sum_{\substack{n \geq 0 \\ m|n}} a_n q^n.$$

Note that if $f(z) = \sum_n a_n q^n$, where $q = e^{2\pi i}$, then we have (for any integer k)

$$V_m f(z) = f(mz) = m^{-k/2} f|_k \beta_m \quad \text{and} \quad U_m f(z) = \tfrac{1}{m} \sum_{j=0}^{m-1} f\left(\tfrac{z+j}{m}\right). \tag{2.30}$$

To determine the effect of T_m on modular forms $f \in M_k(\Gamma_1(N))$, it enough to find an expression for $f|_k T_m$ with $f \in M_k(N, \chi)$ since by Proposition 2.25 every $f \in M_k(\Gamma_1(N))$ is the sum of f_i's with $f_i \in M_k(N, \chi_i)$.

Proposition 2.34 *If $f = \sum a_n(f) q^n \in M_k(N, \chi)$, then the n-th Fourier coefficient of $f|_k T_p$ for a prime p is given by*

$$a_n(f|_k T_p) = a_{pn}(f) + \chi(p) p^{k-1} a_{n/p}(f),$$

where $\chi(p) = 0$ if $p|N$ and $a_{n/p} = 0$ if $p \nmid n$. Thus

$$T_p = U_p + \chi(p) p^{k-1} V_p \quad \text{on } M_k(N, \chi).$$

More generally, for any positive integer m we have

$$a_n(f|_k T_m) = \sum_{d|(m,n)} \chi(d) d^{k-1} a_{mn/d^2}(f), \quad \text{if } n \geq 0, \tag{2.31}$$

and hence

$$T_m = \sum_{d|m} \chi(d) d^{k-1} V_d \circ U_{m/d} \quad \text{on } M_k(N, \chi). \tag{2.32}$$

Proof. [Ko], p. 161–163, or [Sh], equation (3.5.12), p. 80.

Remark 2.35 By comparing the above formula with that of Proposition 1.25, we thus see that in the case of level $N = 1$ the Hecke Operator T_n defined here coincides with the one defined in §1.5.

Hecke observed that many of the interesting modular forms are eigenforms for all the Hecke operators T_n, and that these enjoy some remarkable properties.

Proposition 2.36 *Suppose that $f \in M_k(\Gamma_1(N))$ is an eigenform with respect to all the operators T_n, i.e. $T_n f = \lambda_n f$, for some $\lambda_n \in \mathbb{C}$, for all $n \geq 1$. Then $f \in M_k(N, \chi)$ for some character $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ and we have*

$$a_n(f) = \lambda_n a_1(f), \quad \text{for all } n \geq 1.$$

Thus, $a_1(f) \neq 0$ unless $f = c$ is a constant function.

Proof. By hypothesis, f is an eigenform under T_p and under T_p^2, and hence by Proposition 2.32(c) f is also an eigenform under $T_{p,p}$, if $p \nmid N$ is a prime. Thus, by (2.29), f is a eigenform under the σ_p's, for all primes $p \nmid N$. By Dirichlet's theorem on primes in arithmetic progressions, the σ_p's generate $\Gamma_0(N)/\Gamma_1(N)$, and so f is an eigenform with respect to $\Gamma_0(N)/\Gamma_1(N)$. But this means that $f \in M_k(N, \chi)$, for some character χ. This proves the first assertion, and the other follows easily because $\lambda_n a_1(f) = a_1(\lambda_n f) = a_1(f|_k T_n) \overset{(2.31)}{=} a_n(f)$.

Remark 2.37 If f is a T_n-eigenfunction with $a_0(f) \neq 0$, then the eigenvalue λ_n is completely determined by the character χ (and by n), for we have $\lambda_n = \sum_{d|n} \chi(d) d^{k-1}$; cf. [Ko], p. 163.

Example 2.38 (a) Recall from Example 1.27 that the Eisenstein series E_k and the discriminant form Δ are eigenforms of level 1.

(b) By the same argument as in Example 1.27, we see from Example 2.30(a) that $g(z) = \eta(z)\eta(11z) \in S_2(\Gamma_0(11))$ is a T_n-eigenform for all n because $\dim_{\mathbb{C}} S_2(\Gamma_0(11)) = 1$. More generally, for any $k|24$, $k \equiv 0\,(2)$, we have that $S_k(\Gamma_0(N-1)) = \mathbb{C}g_k$, where $N = \frac{24}{k}$ and $g_k(z) = (\eta(z)\eta((N-1)z))^{k/2}$, and hence g_k is a T_n-eigenform for all $n \geq 1$.

As in the case of level 1, the Fourier coefficients of the q-expansion $f(z) = \sum a_n(f) q^n$ of an *eigenform* $f \in M_k(\Gamma_1(N))$ satisfy some rather remarkable identities, which are best understood in terms of its associated *Dirichlet series* (or *L-function*):

$$L(f, s) := \sum_{n \geq 1} a_n(f) n^{-s}, \quad \text{if } f(z) = \sum_{n \geq 0} a_n(f) q^n, \text{ where } q = e^{2\pi i z}.$$

Corollary 2.39 *Suppose that $f \in M_k(N, \chi)$ is an eigenform with respect to all the Hecke operators T_n. If f is* normalized, *i.e. if $a_1(f) = 1$, then its associated L-function $L(f, s)$ has an Euler product of the form*

$$L(f, s) := \sum_{n \geq 1} a_n(f) n^{-s} = \prod_p (1 - a_p(f) p^{-s} + \chi(p) p^{k-1-s})^{-1}.$$

Conversely, if $f \in M_k(N, \chi)$ is such that its Dirichlet series $L(f, s)$ has an Euler product of this form (for some $a_p \in \mathbb{C}$), then f is a (normalized) eigenform with respect to all T_n's, and for each prime p we have $a_p = a_p(f)$.

Proof. Similar to Proposition 1.32; cf. [Ko], p. 163 or [Mi], p. 149.

Remark 2.40 In Theorem 1.35 we learned that the L-function $L(f,s)$ of a modular form f of level $N=1$ has an Euler product if and only if it has an Euler product of the above type. This, however, is no longer true for higher level N because the Euler factors at the primes $p|N$ may be quite complicated. Nevertheless, an analogous result does hold for the Euler factors at the primes $p \nmid N$; cf. Hecke[He], Satz 42.

For level 1 we found that the normalized T_n-eigenfunctions form a basis of $M_k(\Gamma(1))$; cf. Theorem 1.39. This is no longer true for higher level, as can be seen by examples. However, we do have the following (partial) generalization:

Theorem 2.41 *Let $\mathbb{T}' \subset \mathrm{End}(S_k(\Gamma_1(N)))$ denote the $\mathbb{C}$-algebra generated by all Hecke operators T_n with $(n,N)=1$. Then $S_k(\Gamma_1(N))$ has a basis consisting of $\mathbb{T}'$-eigenforms.*

The proof of this result is very similar to that of Theorem 1.39: we observe that if $(n,N)=1$, then the Hecke operator T_n commutes with its adjoint T_n^* which defined via the Petersson scalar product on $S_k(\Gamma)$:

Notation. If $f,g \in S_k(\Gamma)$, then

$$\langle f,g\rangle_\Gamma = \int_{\Gamma\backslash\mathfrak{H}} f(z)\overline{g(z)}y^{k-2}\,dx\,dy \tag{2.33}$$

is a (positive definite) hermitian pairing on $S_k(\Gamma)$ called the *Petersson pairing*.

Remark 2.42 (a) If $k=2$, then via the identification ω_Γ of (the proof of) Proposition 2.28, this pairing coincides (up to a constant) with the usual hermitian pairing $(\omega_1,\omega_2) = \int_X \omega_1 * \bar{\omega}_2$ on $\Omega^1(X)$ of the compact Riemann surface $X = X_\Gamma$; cf. Springer [Sp], p. 181.

(b) The above integral (2.33) still converges if we allow $f \in M_k(\Gamma)$ (but still require $g \in S_k(\Gamma)$), and so the orthogonal complement $S_k(\Gamma)^\perp = \{f \in M_k(\Gamma) : \langle f,g\rangle = 0, \forall g \in S_k(\Gamma)\}$ can be defined; cf. [Mi], p. 44. It can be shown that this space is generated by Eisenstein series when $k \geq 3$; cf. [Mi], p. 69.

Proposition 2.43 *(a) If $\Gamma_2 = \alpha^{-1}\Gamma_1\alpha$ where $\alpha \in \mathrm{GL}_2(\mathbb{Q})$ and $f_i \in S_k(\Gamma_i)$, then*

$$\langle f_1, f_2|_k\alpha^{-1}\rangle_{\Gamma_1} = \langle f_1|_k\alpha, f_2\rangle_{\Gamma_2}. \tag{2.34}$$

(b) If $\Gamma_1 \leq \Gamma_2$ and $f_i \in S_k(\Gamma_i)$, then

$$\langle f_1, f_2\rangle_{\Gamma_1} = \langle tr_{\Gamma_1/\Gamma_2}(f_1), f_2\rangle_{\Gamma_2}. \tag{2.35}$$

Proof. See [Sh], p. 75 or [Ko], p. 171. (Note that [Ko] defines the Petersson scalar product slightly differently.)

Corollary 2.44 *Let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ and put $\alpha^* = \det(\alpha)\alpha^{-1}$. If Γ_1 and Γ_2 are two congruence subgroups, then*

$$\langle f_1|_k T_{\Gamma_1,\Gamma_2}(\alpha), f_2\rangle_{\Gamma_2} = \langle f_1, f_2|_k T_{\Gamma_2,\Gamma_1}(\alpha^*)\rangle_{\Gamma_1}, \quad \forall f_i \in S_k(\Gamma_i),\ i = 1, 2.$$

Thus, $T_{\Gamma_2,\Gamma_1}(\alpha^*)$ is the adjoint *of $T_{\Gamma_1,\Gamma_2}(\alpha)$ with respect to the Petersson product.*

Proof. Put $\Gamma_\alpha = \Gamma_2 \cap \alpha^{-1}\Gamma_1\alpha$ and $\Gamma_{\alpha^{-1}} = \alpha\Gamma_\alpha\alpha^{-1} = \Gamma_1 \cap \alpha\Gamma_2\alpha^{-1}$. Moreover, put $c = \det(\alpha)^{k/2-1}$. Then by (2.35) and (2.34) we obtain

$$\begin{aligned} c^{-1}\langle f_1|_k T_{\Gamma_1,\Gamma_2}(\alpha), f_2\rangle_{\Gamma_2} &= \langle tr_{\Gamma_\alpha/\Gamma_2} f_1|_k\alpha, f_2\rangle_{\Gamma_2} = \langle f_1|_k\alpha, f_2\rangle_{\Gamma_\alpha} = \langle f_1, f_2|\alpha^{-1}\rangle_{\Gamma_{\alpha^{-1}}} \\ &= \langle f_1, tr_{\Gamma_{\alpha^{-1}}/\Gamma_1}(f_2|_k\alpha^{-1})\rangle_{\Gamma_1} = c^{-1}\langle f_1, f_2|_k T_{\Gamma_2,\Gamma_1}(\alpha^*)\rangle_{\Gamma_1}, \end{aligned}$$

which proves the assertion.

We are now ready to prove Theorem 2.41. For this, we shall prove the following slightly more precise result.

Proposition 2.45 *If $(n, N) = 1$, then the adjoint T_n^* of T_n on $S_k(\Gamma_1(N))$ is $\sigma_n^{-1}T_n$, i.e. we have*

$$\langle f|_k T_n, g\rangle = \langle f, g|\sigma_n^{-1}T_n\rangle, \quad \textit{for all } f, g \in S_k(\Gamma_1(N)).$$

Thus, the algebra $\mathbb{T}'$ is $*$-closed (i.e. $T \in \mathbb{T}' \Rightarrow T^* \in \mathbb{T}'$) and hence $S_k(\Gamma_1(N))$ has a basis consisting of $\mathbb{T}'$-eigenforms.

Proof. Since $\alpha_n^* = n\alpha_n^{-1} = \left(\begin{smallmatrix} n & 0 \\ 0 & 1 \end{smallmatrix}\right) \equiv \sigma_n^{-1}\alpha_n \pmod N$, we see that $\alpha_n^* = \sigma_n^{-1}\alpha_n\gamma$, for some $\gamma \in \Gamma(N)$, and so by Corollary 2.44 we have $T_{1,n}^* = T(\alpha_n^*) = T(\sigma_n^{-1}\alpha_n) = T(\sigma_n^{-1})T_{1,n}$, the latter by Remark 2.31(b), property 3'). Thus, by Proposition 2.32(f) we obtain

$$T_n^* = \sum_{d^2|n} T_{1,n/d^2}^* T_{d,d}^* = \sum_{d^2|N} T(d\sigma_d^{-1})T(\sigma_{n/d^2}^{-1})T_{1,n/d^2} = T(\sigma_n^{-1})\sum_{d^2|N} T(d\sigma_d)T_{1,n/d^2} = \sigma_n^{-1}T_n,$$

where we used the obvious facts that $T_{d,d}^* = T((d\sigma_d)^*) = T(d\sigma_d^{-1})$ and that $T(d\sigma_d^{-1})T(\sigma_{n/d^2}^{-1})$ $= T(d\sigma_n^{-1}\sigma_d) = T(\sigma_n^{-1})T(d\sigma_d)$.

This proves the first assertion. Furthermore, since $\sigma_n \in \mathbb{T}'$ for all $(n, N) = 1$ (cf. proof of Proposition 2.36), and since $\sigma_n^{-1} = \sigma_{n^*}$, where $n^*n \equiv 1 \pmod N$, we see that $T_n^* \in \mathbb{T}'$, $\forall (n, N) = 1$. Thus, $\mathbb{T}'$ is a commutative, $*$-closed algebra, and hence by linear algebra $S_k(\Gamma_1(N))$ has a basis of $\mathbb{T}'$-eigenforms; cf. §1.5.3.

Remark 2.46 (a) If we specialize the above result to $N = 1$, then we obtain Theorem 1.39. Note, however, that while the Hecke operators T_n for level 1 are self-adjoint (cf. Proposition 1.38), this is no longer true for higher level.

(b) As we shall see in the next section, $S_k(N, \chi)$ does not have in general a basis consisting of $\mathbb{T}$-eigenforms; i.e. the algebra $\mathbb{T} \subset \mathrm{End}_{\mathbb{C}}(S_k(\Gamma_1(N)))$ generated by all the Hecke operators T_n, $n \geq 1$ is not semi-simple.

(c) For any $n \geq 1$, the adjoint of T_n on $S_k(\Gamma_1(N))$ is given by

$$T_n^* = w_N T_n w_N^{-1}, \quad \text{where } w_N = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}. \tag{2.36}$$

Indeed, since $w_N \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) w_N^{-1} = \left(\begin{smallmatrix} d & -c/N \\ -bN & a \end{smallmatrix}\right)$, we see that $w_N \alpha_n w_N^{-1} = \alpha_n^*$ and that w_N normalizes $\Gamma_1(N)$. Thus, the same argument as in the proof of Proposition 2.45 shows that (2.36) holds.

Remark 2.47 For simplicity, we had restricted the above discussion of Hecke operators to the case that $\Gamma = \Gamma_1(N)$; note that this also includes the case $\Gamma = \Gamma_0(N)$ because $M_k(\Gamma_0(N)) = M_k(N, 1)$ (trivial Nebentypus character). In Shimura's book [Sh], however, one finds a more general treatment of Hecke operators which applies to all congruence subgroups $\Gamma \geq \Gamma(N)$ which can be conjugated by $\beta_t = \left(\begin{smallmatrix} t & 0 \\ 0 & 1 \end{smallmatrix}\right)$ to a group containing $\Gamma_1(Nt)$, for some $t|N$. For these groups, the study of Hecke operators can be reduced to the corresponding study on $\Gamma_1(Nt)$; cf. [Sh], p. 87.

For example, in the case of $\Gamma = \Gamma(N)$, we see that

$$\beta_N \Gamma(N) \beta_N^{-1} = \Gamma_N := \{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1(N) : c \equiv 0 (N^2) \} \ \geq \ \Gamma_1(N^2)$$

because $\beta_N \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \beta_N^{-1} = \left(\begin{smallmatrix} a & b/N \\ cN & d \end{smallmatrix}\right)$, and so the map $\beta_N^* : f \mapsto f|_k \beta_N$ identifies $M_k(\Gamma(N))$ with the subspace $M_k(\Gamma_N)$ of $M_k(\Gamma_1(N^2))$. This latter subspace can be decomposed as a sum of certain Nebentypus spaces of level N^2; in fact, we have that

$$(M_k(N))|_k \beta_N = M_k(\Gamma_N) = \bigoplus_{\chi} M_k(N^2, \tilde{\chi}) \subset M_k(\Gamma_1(N^2)), \tag{2.37}$$

where the sum runs over all Dirichlet characters χ mod N and $\tilde{\chi}$ denotes the lift of χ to a character mod N^2. (This decomposition is easily verified by observing that the map β_N^* induces for each Dirichlet character χ mod N a bijection

$$\beta_N^* : M_k(\Gamma(N), \chi) \xrightarrow{\sim} M_k(N^2, \tilde{\chi}),$$

where $M_k(\Gamma(N), \chi) = \{ f \in S_k(\Gamma(N)) : f|_k \sigma_a = \chi(a) f, \forall a \in Z_N \}$ denotes the χ-eigenspace of $M_k(\Gamma(N))$ under the natural action of $Z_N := (\mathbb{Z}/N\mathbb{Z})^\times$ as a group of automorphisms of $S_k(\Gamma(N))$ via $a \mapsto \sigma_a$.)

Therefore, by the above identification (2.37) we can then transport the theory of Hecke operators to $\Gamma(N)$. For example, we define the Hecke operator $T_n^{\Gamma(N)}$ on $M_k(\Gamma(N))$ by

$$f|_k T_n^{\Gamma(N)} = ((f|_k \beta_N)|_k T_n)|_k \beta_N^{-1}, \text{ if } f \in M_k(\Gamma(N)).$$

It then immediate that all the corresponding properties hold for the $T_n^{\Gamma(N)}$'s.

2.4 Atkin-Lehner Theory

In the previous section we saw that the $\mathbb{T}$-eigenfunctions $f \in S_k(\Gamma_1(N))$ have many interesting properties and raised the question of the *existence* of such functions. In the case of level $N = 1$ we had already obtained a complete answer to this: the normalized $\mathbb{T}$-eigenfunction of $S_k(\Gamma(1))$ form a basis of the space; cf. Theorem 1.39 or Proposition 2.34. For higher level, however, this existence question is much more subtle, as we shall see. Let us briefly review what we know so far:

1) If $f \in \mathbb{V} := S_k(\Gamma_1(N))$ is a $\mathbb{T}$-eigenfunction, then the associated $\mathbb{T}$-eigenspace $\mathbb{V}_{\chi_f}$ is 1-dimensional and contains a unique normalized $\mathbb{T}$-eigenfunction; cf. Proposition 2.36. (Here, as in (1.70), $\chi_f : \mathbb{T} \to \mathbb{C}$ is the character defined by $f|_k T = \chi_f(T) f$, $\forall T \in \mathbb{T}$.) Thus

$$\#\{f \in S_k(\Gamma_1(N)) : f \text{ is a normalized } \mathbb{T}\text{-eigenfunction}\} = \#\hat{\mathbb{T}}.$$

   But in general $\mathbb{V}$ does not have a basis consisting of $\mathbb{T}$-eigenfunctions, i.e. $\#\hat{\mathbb{T}} < \dim \mathbb{V}$, as as Example 2.51 below shows.

2) By Proposition 2.36 we know that $\mathbb{V} := S_k(\Gamma_1(N))$ does have a basis consisting of $\mathbb{T}'$-eigenfunctions, where $\mathbb{T}' = \langle T_n : n \geq 1, (n, N) = 1\rangle \subset \mathrm{End}_{\mathbb{C}}(\mathbb{V})$ is the $\mathbb{C}$-algebra generated by the Hecke operators T_n prime to the level. Thus we have

$$\mathbb{V} = \bigoplus_{\chi' \in \hat{\mathbb{T}}'} \mathbb{V}_{\chi'}.$$

   However, in general the eigenspaces $\mathbb{V}_{\chi'} = \{f \in \mathbb{V} : f|_k T = \chi'(T) f, \forall T \in \mathbb{T}'\}$ are not 1-dimensional.

This, therefore, raises the following problems and questions.

Problem 1. Describe the set of characters or, equivalently, describe the set of $\mathbb{T}$-eigenfunctions of $\mathbb{V} = S_k(\Gamma_1(N))$.

Problem 2. Describe explicitly the above decomposition of $\mathbb{V}$ into its $\mathbb{T}'$-eigenspaces. More precisely:

(a) Describe the set $\hat{\mathbb{T}}'$ of characters of $\mathbb{T}$';

(b) For each character $\chi' \in \hat{\mathbb{T}}'$, determine a basis for the $\mathbb{T}'$-eigenspace $\mathbb{V}_{\chi'}$.

In 1970 A.O.L. Atkin and J. Lehner [AL] made the following remarkable discovery. While it is not possible to find a basis of eigenforms for the whole of $S_k(\Gamma_1(N))$, one can define a certain subspace $S_k^{\mathrm{new}}(\Gamma_1(N)) \subset S_k(\Gamma_1(N))$ which does have such a basis, and which has a natural complement $S_1^{\mathrm{old}}(\Gamma_1(N))$ which consists of the forms "coming from lower level". (Actually, Atkin-Lehner only considered the subspace $S_k(\Gamma_0(N))$, and the extension to $S_k(\Gamma_1(N))$ was later worked out by Miyake[Mi1] and Li[Li].) As a result, one obtains: (i) a *complete* answer to Problem 2; (ii) a *satisfactory* answer to Problem 1; and (iii) a "canonical basis" of $S_k(\Gamma_1(N))$, of which only a part consists of $\mathbb{T}$-eigenforms.

2.4.1 The Definition of Newforms

We begin by defining *oldforms* on $\Gamma_1(N)$; these are modular forms which come from lower level by "twisting" by the operator β_d. For this, we first note that for any positive divisors $M|N$ and $d|\frac{N}{M}$ we have $\Gamma_1(N) \leq \beta_d^{-1}\Gamma_1(M)\beta_d$ (cf. (2.3)), and hence the rule

$$f \;\mapsto\; f|_k T_{\Gamma_1(M),\Gamma_1(N)}(\beta_d) \;=\; d^{k/2-1} f|_k\beta_d \;=\; d^{k-1}\sum a_n(f)q^{nd} \tag{2.38}$$

defines an injective linear map

$$\beta_{M,d}^{(N)} = T_{\Gamma_1(M),\Gamma_1(N)}(\beta_d) : S_k(\Gamma_1(M)) \to S_k(\Gamma_1(N)).$$

Definition. The *space of oldforms* or *old subspace* is the space spanned by all forms $f|_k\beta_d$ for $dM|N$, $M \neq N$ and f in $S_k(\Gamma_1(M))$:

$$S_k^{\text{old}}(\Gamma_1(N)) := \sum_{dM|N, M\neq N} S_k(\Gamma_1(M))|_k\beta_{M,d}^{(N)}.$$

The *space of newforms* or *new subspace* is the orthogonal complement of the old subspace with respect to the Petersson inner product,

$$S_k^{\text{new}}(\Gamma_1(N)) = S_k^{\text{old}}(\Gamma_1(N))^{\perp}.$$

Thus

$$S_k(\Gamma_1(N)) = S_k^{\text{old}}(\Gamma_1(N)) \oplus S_k^{\text{new}}(\Gamma_1(N))$$

As we shall see, this direct sum is a decomposition of $\mathbb{T}$-modules, where $\mathbb{T} = \mathbb{T}_N \subset \text{End}(S_k(\Gamma_1(N)))$ is the *Hecke algebra* of level N, i.e. the $\mathbb{C}$-algebra generated by the Hecke operators $T_n = T_n^{(N)} \in \text{End}(S_k(\Gamma_1(N)))$, for all $n \geq 1$. Note that one has to be careful about the level N, for the actions of the algebras $\mathbb{T}_N$ and $\mathbb{T}_M$ are not completely compatible (cf. (2.44) below). Nevertheless, we do have the following (partial) compatibility relation:

Proposition 2.48 *If $dM|N$ and n is an integer which is coprime to N, then the following diagram commutes:*

$$\begin{array}{ccc} S_k(\Gamma_1(N)) & \xrightarrow{T_n^{(N)}} & S_k(\Gamma_1(N)) \\ {\scriptstyle \beta_{M,d}}\uparrow & & \uparrow{\scriptstyle \beta_{M,d}} \\ S_k(\Gamma_1(M)) & \xrightarrow{T_n^{(M)}} & S_k(\Gamma_1(M)) \end{array} \tag{2.39}$$

Proof. Let $f \in S_k(M,\chi)$. Then $f|_k\beta_d \in S_k(N,\tilde{\chi})$, where $\tilde{\chi}$ is the lift of $\chi : \mathbb{Z}/M\mathbb{Z} \to \mathbb{C}^\times$ to $\mathbb{Z}/N\mathbb{Z}$, and so, by (2.32) (and (2.30)) we have

$$f|_k T_n^{(M)}|_k\beta_d = \sum_{t|n} \chi(t)t^{k-1}d^{\frac{k}{2}}V_dV_tU_{n/t}(f) \text{ and } f|_k\beta_d|_k T_n^{(N)} = \sum_{t|n} \tilde{\chi}(t)t^{k-1}d^{\frac{k}{2}}V_tU_{n/t}V_d(f).$$

Now since $(d,t) = 1$ for all $t|n$, we have that $U_tV_d = V_dU_t$ (and also $V_tV_d = V_dV_t$, as well as $\tilde{\chi}(t) = \chi(t)$), and so $f|_k T_n^{(M)}|_k\beta_d = f|_k\beta_d|_k T_n^{(N)}$, from which the assertion follows.

Remark 2.49 (a) It follows from (2.39) that $S_k^{\rm old}(\Gamma_1(N))$ is stable under the Hecke algebra $\mathbb{T} = \mathbb{T}'_N$ generated by Hecke operators $T_n^{(N)}$, for $(n,N)=1$, and hence the same is true for $S_k^{\rm new}(\Gamma_1(N))$ because $\mathbb{T}'_N$ is $*$-closed (cf. Proposition 2.45). In fact, as we shall see later (cf. Remark 2.57(c)), both $S_k^{\rm old}$ and $S_k^{\rm new}$ are stable under the full Hecke algebra $\mathbb{T}_N$, but this fact is more subtle.

(b) If we take $d=1$ in the above proposition, then (2.39) shows that $T_n^{(M)}$ is the restriction of $T_n^{(N)}$ to $S_k(\Gamma_1(M)) \subset S_k(\Gamma_1(N))$, *provided that* $(n,N)=1$. (If $(n, \frac{N}{M}) > 1$, then this assertion is often *false*, as equation (2.44) below shows by taking $n = p|N$.)

Note that the above is an *analytic* definition of the space of newforms (since it uses the Petersson product). However, this space can also be defined *algebraically*, either in terms of the algebra $\mathbb{T}'$ as in Corollary 2.55 or, more directly, by using the *degeneracy operators* $D_{M,d}$ and $D_{M,d}^* \in \mathrm{End}_{\mathbb{C}}(S_k(\Gamma_1(N))$ which are defined as follows.

Notation. Let $M, d \geq 1$ be integers such that $dM|N$, and let $f \in S_k(\Gamma_1(N))$. Put

$$f|_k D_{M,d} = d^{k/2-1}(tr_{\Gamma_1(N)/\Gamma_1(M)}(f))|_k\beta_d, \quad f|_k D_{M,d}^* = d^{k/2-1}(tr_{\Gamma_1(\frac{N}{d},d)/\Gamma_1(M)}(f))|_k\alpha_d,$$

where $\Gamma_1(\frac{N}{d},d) = \alpha_d^{-1}\Gamma_1(N)\alpha_d = \{\left(\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) : x \equiv w \equiv 1\ (N), d|y, \frac{N}{d}|z\}$.

Proposition 2.50 *We have*

$$(2.40)\quad S_k(\Gamma_1(N))^{\rm old} = \sum_{\substack{dM|N \\ M \neq N}} Im(D_{M,d}) \quad \textit{and} \quad S_k(\Gamma_1(N))^{\rm new} = \bigcap_{\substack{dM|N \\ M \neq N}} Ker(D_{M,d}^*).$$

Proof. We first observe that

$$(2.41)\qquad D_{M,d} = \beta_{M,d} \circ (\beta_{M,1})^* \quad \text{and} \quad D_{M,d}^* = \beta_{M,1} \circ (\beta_{M,d})^*$$

where $(\beta_{M,d})^*$ denotes the adjoint of $\beta_{M,d} = \beta_{M,d}^{(N)}$ with respect to the Petersson product. Indeed, since $\beta_{M,1} : S_k(\Gamma_1(M)) \to S_k(\Gamma_1(N))$ is just the canonical injection, we have by (2.35) that $\beta_{M,1}^* = tr_{\Gamma_1(N)/\Gamma_1(M)}$, and so the first formula of (2.41) is clear. Moreover, since $\beta_{M,d}^* = T(\beta^*) = T_{\Gamma_1(N),\Gamma_1(M)}(\alpha_d)$ by Corollary 2.44 and since $\Gamma_1(M) \cap \alpha_d^{-1}\Gamma_1(N)\alpha_d = \Gamma_1(M) \cap \Gamma_1(\frac{N}{d},d) = \Gamma_1(\frac{N}{d},d)$ because $M|\frac{N}{d}$, the second formula of (2.41) follows.

Since $\mathrm{Im}((\beta_{M,1})^*) = S_k(\Gamma_1(M))$ (because $tr(g) = [\Gamma_1(M) : \Gamma_1(N)]g$, for $g \in S_k(G_1(M))$), the first formula of (2.41) shows that $\mathrm{Im}(D_{M,d}) = \mathrm{Im}(\beta_{M,d})$, and so the first equation of (2.40) is clear. Moreover, since (2.41) shows that $D_{M,d}^*$ is the adjoint of $D_{M,d}$, we obtain:

$$\begin{aligned} f \in S_k(\Gamma_1(N))^{\rm new} &\Leftrightarrow \langle f, g|_k D_{M,d}\rangle = 0, \quad \forall g \in S_k(\Gamma_1(N)),\ \forall dM|N, M \neq N, \\ &\Leftrightarrow \langle f|_k D_{M,d}^*, g\rangle = 0, \quad \forall g \in S_k(\Gamma_1(N)),\ \forall dM|N, M \neq N, \\ &\Leftrightarrow f|_k D_{M,d}^* = 0, \quad \forall dM|N, M \neq N, \end{aligned}$$

which proves the second formula of (2.40).

Before stating the main theorems of Atkin-Lehner theory, it is perhaps useful to work out the following example which illustrates the basic difficulty with the action of Hecke operators on oldforms.

Example 2.51 Let $f \in S_k(\Gamma_0(M)) = S_k(M,1)$ be a normalized $\mathbb{T}_M$-eigenform and suppose that $p \nmid M$ is a prime. Fix $r \geq 3$ and put $N = p^r M$. Then the space

$$S_f = \langle f(z), f(pz), f(p^2 z), \ldots, f(p^r z) \rangle \subset S_k^{\text{old}}(\Gamma_0(N))$$

is stable under the Hecke algebra $\mathbb{T}_N \subset \text{End}(S_k(\Gamma_0(N)))$ but S_f does *not* have a basis of eigenforms under $\mathbb{T}_N$.

To see this, write $f_j(z) = f(p^j z) = V_{p^j} f = p^{-jk/2} f|_k \beta_{p^j}$, for $0 \leq j \leq r$. If $(n, N) = 1$, then by (2.39) we have

$$f_j|_k T_n^{(N)} = p^{-jk/2} f|_k T_n^{(M)} \beta_{p^j} = p^{-jk/2} a_n(f) f|_k \beta_{p^j} = a_n(f) f_j, \quad \text{for } 0 \leq j \leq r. \tag{2.42}$$

Furthermore, by (2.32) we have $T_p^{(N)} = U_p$, so

$$T_p^{(N)} f_j = f_{j-1}, \quad \text{for } 1 \leq j \leq r. \tag{2.43}$$

On the other hand, since f is an eigenfunction of $T_p^{(M)}$, we have $f|_k T_p^{(M)} = a_p f$, where $a_p = a_p(f)$. Now by (2.32) we have $f|_k T_p^{(M)} = U_p f + p^{k-1} V_p f = f_k T^{(N)} + p^{k-1} f_1$. Thus we obtain

$$f|_k T_p^{(N)} = f|_k T_p^{(M)} - p^{k-1} f_1 = a_p f - p^{k-1} f_1; \tag{2.44}$$

in particular, $f|_k T_p^{(N)} \neq f|_k T_p^{(M)}$ and f is no longer an eigenfunction with respect to $T_p^{(N)}$.

From equations (2.42)–(2.44) we see that S_f is a $\mathbb{T}_N$-submodule, and that the matrix of $T_p^{(N)}$ with respect to the basis $f_0 = f, f_1, \ldots f_r$ is

$$\begin{pmatrix} a_p & 1 & 0 & \ldots & \ldots & 0 \\ -p^{k-1} & 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 1 & \ldots & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & \ldots & 0 & 1 \\ 0 & \ldots & \ldots & \ldots & 0 & 0 \end{pmatrix}.$$

Thus, the characteristic polynomial of $T_p^{(N)}$ on S_f is $\text{ch}(x) = x^{r-1}(x^2 - a_p x + p^{k-1}) = x^{r-1}(x-\alpha)(x-\beta)$ and so the $T_p^{(N)}$-eigenfunctions are precisely the scalar multiples of the functions $f_0 - a_p f_1 + p^{k-1} f_2$, $f_0 - \beta f_1$ and $f_0 - \alpha f_1$ (with eigenvalues 0, α and β, respectively). We thus see that if $r \geq 3$, then $T_p^{(N)}$ is not diagonalizable on S_f, i.e. S_f does not have a basis of $T_p^{(N)}$-eigenfunctions.

2.4.2 Basic Results

In this section we shall present the main results of Atkin-Lehner Theory which give a satisfactory answer to the basic questions raised at the beginning of this section. As we shall see in the next subsection, most of these results are direct consequences of the Main Theorem 2.58 of Atkin-Lehner Theory which will presented later.

Since in the sequel we shall frequently deal with modular forms of (fixed) weight k with varying level N, it is convenient to introduce the abbreviation

$$\mathbb{V}_N = S_k(\Gamma_1(N)).$$

Definition. If $f \in \mathbb{V}_N^{\text{new}}$ is a $\mathbb{T}'$-eigenform with $a_1(f) = 1$, then we call f a *normalized newform* of level N. The set of all normalized newforms in $\mathbb{V}_N$ is denoted by $\mathcal{N}(\mathbb{V}_N)$.

Theorem 2.52 *Each $f \in \mathcal{N}(\mathbb{V}_N)$ is a $\mathbb{T}$-eigenfunction, and hence $\mathcal{N}(\mathbb{V}_N)$ is a basis of $\mathbb{V}_N^{\text{new}}$ consisting of $\mathbb{T}$-eigenfunctions. Thus $\mathbb{V}_N^{\text{new}}$ is a $\mathbb{T}$-module and $\#\mathcal{N}(\mathbb{V}_N) = \dim V_N^{\text{new}}$.*

Thus, we see that we have a rich supply of $\mathbb{T}$-eigenfunctions. While this result doesn't classify all the $\mathbb{T}$-eigenfunctions, it gives a satisfactory answer to Problem 1 above in the sense that it classifies all the eigenfunctions which do not come from lower level.

We now turn to Problem 2, i.e. the classification of the $\mathbb{T}'$-eigenfunctions. For this, let us introduce the following notation.

Notation. For any $M|N$ and $f \in \mathcal{N}(\mathbb{V}_M)$, let

$$S_f(N) \;=\; S_f(\mathbb{V}_N) \;=\; \sum_{d|N/M} \mathbb{C} f|_k\beta_d \;=\; \bigoplus_{d|N/M} \mathbb{C} f|_k\beta_d;$$

clearly $\dim S_f(N) = \sigma_0(\frac{N}{M})$ = number of divisors of $\frac{N}{M}$. Furthermore, we let

$$\mathcal{N}^*(\mathbb{V}_N) := \bigcup_{M|N} \mathcal{N}(\mathbb{V}_M)$$

denote the set of *normalized newforms of all levels $M|N$*.

It is clear from the definition and Proposition 2.48 that every $f \in \mathcal{N}^*(\mathbb{V}_N)$ is a $\mathbb{T}'$-eigenfunction. If $\chi'_f \in \hat{\mathbb{T}}'_N$ denotes the associated character, then is clear from Proposition 2.48 that every $g \in S_f(N)$ is a χ'_f-eigenfunction, i.e. that $S_f(N) \subset (\mathbb{V}_N)_{\chi'_f}$. We now have:

Theorem 2.53 (Atkin-Lehner Decomposition) *For each $f \in \mathcal{N}^*(\mathbb{V}_N)$, the space $S_f(N)$ is the χ'_f-eigenspace of $\mathbb{V}_N$, i.e.*

$$S_f(N) = (\mathbb{V}_N)_{\chi'_f},$$

and hence $S_f(N)$ is also a $\mathbb{T}$-module. Furthermore, the map $f \mapsto \chi'_f$ induces a bijection $\mathcal{N}^*(\mathbb{V}_N) \xrightarrow{\sim} \hat{\mathbb{T}}'_N$, and hence we have the $\mathbb{T}$-module decomposition

$$\mathbb{V}_N \;=\; \bigoplus_{M|N} \bigoplus_{f \in \mathcal{N}(\mathbb{V}_M)} S_f(N) \;=\; \bigoplus_{f \in \mathcal{N}^*(\mathbb{V}_N)} S_f(N). \tag{2.45}$$

Remark 2.54 (a) The decomposition (2.45) shows that the set

$$\mathcal{B}(\mathbb{V}_N) := \{f|_k\beta_d : f \in \mathcal{N}(\mathbb{V}_M),\ M|N,\ d|\tfrac{N}{M}\}$$

is a *basis* of $\mathbb{V}_N$. Thus, $\mathbb{V}_N$ has a "canonical basis" consisting of normalized newforms of all levels, together with certain "twists" of these with respect to the operators β_d, $d|N$.

(b) Note that in general not every function in $\mathcal{B}(\mathbb{V}_N)$ is a $\mathbb{T}_N$-eigenfunction. For example, $f|\beta_d$ can never be a $\mathbb{T}_N$-eigenfunction if $d \neq 1$ because we have $a_1(f|\beta_d) = 0$ if $d \neq 1$. Moreover, even if $d = 1$ but $f \in \mathcal{N}(\mathbb{V}_M)$, $M \neq N$, then f need not be a $\mathbb{T}_N$-eigenfunction, as is evident in the situation of Example 2.51. On the other hand, if $M \neq N$ has the same prime divisors as N, then $f \in \mathcal{N}(\mathbb{V}_M)$ is a $\mathbb{T}_N$-eigenfunction.

(c) For every $f \in \mathcal{N}^*(\mathbb{V}_N)$, the space $S_f(N)$ contains at least one $\mathbb{T}_N$-eigenfunction; in other words, the restriction map $\chi \mapsto \chi|\mathbb{T}'$ defines a *surjection*

$$\hat{\mathbb{T}}_N \ \to\ \hat{\mathbb{T}}'_N.$$

[To see this, note that first we have a canonical bijection between the set of characters $\hat{\mathbb{T}}$ and the set $\max(\mathbb{T})$ of maximal ideals of $\mathbb{T}$ (via $\chi \mapsto \mathrm{Ker}(\chi)$), and the same is true for $\hat{\mathbb{T}}'$. Since every maximal ideal of $\mathbb{T}'$ is contained in a maximal ideal of $\mathbb{T}$ (by the Going-up Theorem of Commutative Algebra), every character of $\mathbb{T}'$ lifts to a character of $\mathbb{T}$.]

(d) In general, the above map $\hat{\mathbb{T}} \to \hat{\mathbb{T}}'$ is not injective, for there may be several $\mathbb{T}_N$-eigenfunctions contained in a single $\mathbb{T}'_N$-eigenspace $S_f(N)$, as we already saw in Example 2.51.

As an application of the above result, we also obtain the following algebraic characterization of the newspace V_N^{new}.

Corollary 2.55 *The space $\mathbb{V}_N^{\mathrm{new}}$ is the sum of the eigenspaces of T'_N whose eigencharacters occur with multiplicity one, whereas the space $\mathbb{V}_N^{\mathrm{old}}$ is the sum of the eigenspaces whose eigencharacters appear with multiplicity greater than one.*

Proof. By Theorem 2.53 we know that every $\mathbb{T}'$-eigenspace is of the form $S_f(N)$, for some $f \in \mathcal{N}(\mathbb{V}_M)$, $M|N$. Since $\dim S_f(N) = \sigma_0(N/M)$, we see that $\dim S_f(N) = 1 \Leftrightarrow N = M \Leftrightarrow f \in \mathcal{N}(\mathbb{V}_N) \Leftrightarrow S_f(N) \subset \mathbb{V}_N^{\mathrm{new}}$. On the other hand, if $\dim S_f(N) > 1$, then $S_f(N) \subset \mathbb{V}_N^{\mathrm{old}}$, and so the assertion follows from the decomposition (2.45).

Corollary 2.56 *Each $f \in \mathcal{N}(\mathbb{V}_N)$ is a $(\mathbb{T}_N \cup \mathbb{T}_N^*)$-eigenform, where $\mathbb{T}_N^* = \{T^* \in End(\mathbb{V}_N) : T \in \mathbb{T}_N\}$ is the subalgebra consisting of all adjoints of $\mathbb{T}_N$. Thus we have*

$$f|_k T_n^* = \overline{a_n(f)} f, \quad \text{for all } n \geq 1. \tag{2.46}$$

Moreover, if $f^* := \sum_{n\geq 1} \overline{a_n(f)} q^n$, then $f^* \in \mathcal{N}(\mathbb{V}_N)$ and we have

$$f|_k w_N = c f^*, \quad \text{for some } c \in \mathbb{C}^\times. \tag{2.47}$$

In particular, w_N maps $\mathbb{V}_N^{\mathrm{new}}$ (and $\mathbb{V}_N^{\mathrm{old}}$) into itself.

Proof. Since $\mathbb{T}_N(N)$ is $*$-closed, we see that $\mathbb{T}_N(N) \subset \mathbb{T}_N \cap \mathbb{T}_N^*$. Now since $\mathbb{T}_N$ is commutative, so is $\mathbb{T}_N^*$, and hence T^* commutes with $\mathbb{T}_N(N)$, if $T \in \mathbb{T}_N$. Thus, $f|T^*$ is in the $\mathbb{T}_N(N)$-eigenspace of f, and so by the Multiplicity-One Theorem we have $f|T^* = c_T f$, for some $c_T \in \mathbb{C}$. This proves the first statement.

Thus, $f|T_n^* = c_n f$, for some $c_n \in \mathbb{C}$. Since also $f|T_n = a_n(f)f$, we obtain $c_n\langle f, f\rangle = \langle f|T_n^*, f\rangle = \langle f, f|T_n\rangle = \langle f, a_n(f)f\rangle = \overline{a_n(f)}\langle f, f\rangle$, and so $c_n = \overline{a_n(f)}$. Thus, (2.46) holds.

To prove (2.47), we first recall that w_N normalizes $\Gamma_1(N)$; cf. Remark 2.46c). Furthermore, since $\beta_M w_N \beta_N^{-1} = w_M$, $\forall M|N$, we see that w_N maps $\mathbb{V}_{N\text{old}}$ into itself, and hence also $\mathbb{V}_N^{\text{new}} = (\mathbb{V}_N^{\text{old}})^\perp$, because $w_N^* = -w_N$. Thus $g := f|_k w_N \in \mathbb{V}_N^{\text{new}}$. Furthermore, since $T_n^* = w_N T_n w_N^{-1}$ by (2.36), we obtain $g|_k T_n = f|_k w_N T_n = f|T_n^* w_N = \overline{a_n(f)} f|_k w_N = \overline{a_n(f)} g$. This means that g is a $\mathbb{T}$-eigenform with T_n-eigenvalue $\overline{a_n(f)}$, and so (2.47) holds with $c = a_1(g)$; cf. Proposition 2.36. Thus $f^* \in \mathbb{V}_N^{\text{new}}$ is a $\mathbb{T}$-eigenform, and hence $f^* \in \mathcal{N}(\mathbb{V}_N)$.

Remark 2.57 (a) Note that while the algebra $\langle \mathbb{T}_N, \mathbb{T}_N^*\rangle$ generated by $\mathbb{T}_N$ and $\mathbb{T}_N^*$ is $*$-closed, it is not commutative in general, for otherwise $\mathbb{V}_N$ would have a basis consisting of $\mathbb{T}$-eigenforms.

(b) It is immediate from Corollaries 1 and 4 that $\mathbb{V}_N^{\text{new}}$ and $\mathbb{V}_N^{\text{old}}$ are $\mathbb{T}_N$-modules as well as $\mathbb{T}_N^*$-modules.

2.4.3 The Main Theorem

The key for the proof of the Atkin-Lehner theorem is the following Structure Theorem 2.58 which, for a given integer D, analyzes the space

$$\mathbb{V}_N(D) = \{f = \sum a_n q^n \in \mathbb{V}_N : a_n = 0 \text{ for all } n \text{ with } \gcd(n, D) = 1\}.$$

Clearly, $\mathbb{V}_N(D) = \mathbb{V}_N(\text{rad}(D))$, i.e. $\mathbb{V}_N(D)$ only depends on the *radical* $\text{rad}(D) = \prod_{p|D} p$ of D, and we have $\mathbb{V}_N(D) \subset \mathbb{V}_N(D')$, if $D|D'$. Also, we observe that by equations (2.30) and (2.38) we have for any $D \geq 1$

$$\beta_d^* \mathbb{V}_M(D) \subset \mathbb{V}_N(dD), \quad \text{if } dM|N. \tag{2.48}$$

Furthermore, if, as above, $\mathbb{T}_N(D) \subset \mathbb{T}_N \subset \text{End}(\mathbb{V}_N)$ denotes the subalgebra generated by all the Hecke operators $T_n^{(N)}$, for $(n, D) = 1$, then it follows from (2.32) that $\mathbb{V}_N(D')$ is a $\mathbb{T}_N(D)$-module if $D'|D$, i.e.

$$\mathbb{V}_N(D')|T \subset \mathbb{V}_N(D'), \quad \text{if } T \in \mathbb{T}_N(D) \text{ and } D'|D. \tag{2.49}$$

The following theorem is the "Main Theorem" of Atkin-Lehner theory.

Theorem 2.58 (Structure Theorem) *We have $\mathbb{V}_N(ND) = \mathbb{V}_N(N) \subset \mathbb{V}_N^{\text{old}}$, for any $D \geq 1$. More precisely,*

$$\mathbb{V}_N(ND) = \sum_{p|N} \beta_p^* \mathbb{V}_{N/p}(N).$$

Before sketching the proof of this theorem in the next subsection, let us deduce its most important consequences. An immediate corollary is the following.

Corollary 2.59 *If $f \in \mathbb{V}_N$ is a $\mathbb{T}_N(ND)$-eigenform for some $D \geq 1$ and $a_1(f) = 0$, then $f \in \mathbb{V}_N(ND) \subset \mathbb{V}_N^{\mathrm{old}}$. Thus, if $0 \neq f \in \mathbb{V}_N^{\mathrm{new}}$ is a $\mathbb{T}(ND)$-eigenform, then $a_1(f) \neq 0$.*

Proof. If $(n, ND) = 1$, then by (2.31) we have $a_n(f) = \lambda_n a_1(f) = 0$, so $f \in \mathbb{V}_N(ND) \subset \mathbb{V}_N^{\mathrm{old}}$ by the Structure Theorem 2.58.

Definition. If $f \in \mathbb{V}_N^{\mathrm{new}}$ is a $\mathbb{T}(N)$-eigenform with $a_1(f) = 1$, then we call f a *normalized newform* of level N. The set of all normalized newforms in $\mathbb{V}_N$ is denoted by $\mathcal{N}(\mathbb{V}_N)$.

Corollary 2.60 *$\mathbb{V}_N^{\mathrm{new}}$ has a basis consisting of normalized newforms.*

Proof. By Remark 2.49a) we know that $\mathbb{V}_N^{\mathrm{new}}$ is a $\mathbb{T}_N(N)$-module, and so Proposition 2.45 shows that $\mathbb{V}_N^{\mathrm{new}}$ has a basis $f_1, \ldots, f_r$ consisting of $\mathbb{T}(N)$-eigenforms. By Corollary 1 we have $a_1(f_i) \neq 0$, and so if we put $\tilde{f}_i = f_i / a_1(f_i)$, then $\tilde{f}_1, \ldots, \tilde{f}_r$ is a basis consisting of normalized newforms.

In fact, it turns out that $\mathcal{N}(\mathbb{V}_N)$ itself is the unique basis consisting of normalized newforms. This and much more follows from the following fundamental result.

Theorem 2.61 (Multiplicity-One Theorem) *Let $f, g \in \mathbb{V}_N$ be $\mathbb{T}(ND)$-eigenforms, for some $D \geq 1$. If $f \neq 0$ and g have the same eigenvalues, i.e., if*

$$f|_k T = a_T f, \quad g|_k T = a_T g \quad \text{for all } T \in \mathbb{T}(ND),$$

and if $f \in \mathbb{V}_N^{\mathrm{new}}$, then $g = cf$, for some $c \in \mathbb{C}$.

Proof. Since $a_1(f) \neq 0$ by Corollary 1, we may assume without loss of generality that $a_1(f) = 1$.

Write $g = g^{\mathrm{old}} + g^{\mathrm{new}}$. Since $\mathbb{T}(ND)$ preserves the old and new subspaces (cf. Remark 2.49), the equation $g|_k T = ag$, with $T \in \mathbb{T}(ND)$, implies that

$$g^{\mathrm{old}}|_k T = a g^{\mathrm{old}} \text{ and } g^{\mathrm{new}}|_k T = a g^{\mathrm{new}},$$

and hence both g^{old} and g^{new} have the same $\mathbb{T}(ND)$-eigenvalues as f.

Now the form $h = a_1(g) f - g^{\mathrm{new}} \in \mathbb{V}_N^{\mathrm{new}}$ is a $\mathbb{T}(ND)$-eigenform with $a_1(h) = 0$, and so by Corollary 1 (of Theorem 2.58) we have $a_1(g) f - g^{\mathrm{new}} \in \mathbb{V}_N^{\mathrm{old}}$, which implies that $g^{\mathrm{new}} = cf$ with $c = a_1(g)$.

It remains to show that $g^{\mathrm{old}} = 0$, so that $cf = g$, as claimed. For this, note first that it follows (by induction) from the definition of the old and new spaces that we can write

$$\mathbb{V}_N = \sum_{\substack{M|N \\ d|\frac{N}{M}}} \beta_d^* \mathbb{V}_M^{\mathrm{new}} \quad \text{and} \quad \mathbb{V}_N^{\mathrm{old}} = \sum_{\substack{M|N \\ M \neq N \\ d|\frac{N}{M}}} \beta_d^* \mathbb{V}_M^{\mathrm{new}}. \tag{2.50}$$

Thus, we have $g^{\text{old}} = \sum g_i$ with $g_i \in \beta_{d_i}^* \mathbb{V}_{M_i}^{\text{new}}$. Now the space $\mathbb{V}_{M_i}^{\text{new}}$ has a basis of $\mathbb{T}_{M_i}(M_i)$-eigenforms (cf. Corollary 2). Since these are also $\mathbb{T}_N(ND)$-eigenforms, we can express g^{old} in the form $g^{\text{old}} = \sum h_j|_k \beta_{d_j}$, where each $h_j \in \mathbb{V}_{M_j}^{\text{new}}$ is a $\mathbb{T}(ND)$-eigenform with the same eigenvalues as g^{old} and therefore, as f. Then $k_j := a_1(h_j)f - h_j$ is a $\mathbb{T}(ND)$-eigenform with $a_1(k_j) = 0$, and so $k_j \in \mathbb{V}_N^{\text{old}}$ by the above Corollary 1. Thus $a_1(h_j)f \in \mathbb{V}^{\text{new}} \cap \mathbb{V}^{\text{old}} = \{0\}$, so $a_1(h_j) = 0$. But then Corollary 1, applied to $h_j \in \mathbb{V}_{M_j}^{\text{new}}$, shows that $h_j = 0$, and so $g^{\text{old}} = 0$, as claimed.

Corollary 2.62 *If $f \in \mathbb{V}_N^{\text{new}} = S_k^{\text{new}}(\Gamma_1(N))$ is a $\mathbb{T}(ND)$-eigenform, then it is an eigenform under the full Hecke algebra $\mathbb{T}$, and $f = cg$, for some $g \in \mathcal{N}(\mathbb{V}_N)$. In particular, $\mathcal{N}(\mathbb{V}_N)$ is a basis of $\mathbb{V}_N^{\text{new}}$ consisting of $\mathbb{T}$-eigenforms, and hence $\mathbb{V}_N^{\text{new}}$ is a $\mathbb{T}$-module.*

Proof. By the Multiplicity-One Theorem 2.61, the $\mathbb{T}(ND)$-eigenspace of f has dimension one. Now for any $T \in \mathbb{T}$, $f|T$ is a $\mathbb{T}(ND)$-eigenfunction in the same eigenspace as f, since $\mathbb{T}$ is a commutative algebra. Thus, $f|T$ must be a constant multiple of f, i.e. $f|T = cf$, so f is a $\mathbb{T}_N$-eigenform. Moreover, $a_1(f) \neq 0$ by the above Corollary 1 (of Theorem 2.58), and so $g = f/a_1(f) \in \mathcal{N}(\mathbb{V}_N)$.

Now by Theorem 2.61 again, any two elements of $\mathcal{N}(\mathbb{V}_N)$ belong to different $\mathbb{T}(N)$-eigencharacters, and so $\mathcal{N}(\mathbb{V}_N)$ is a linearly independent set. Thus, $\mathcal{N}(\mathbb{V}_N)$ is a basis of $\mathbb{V}_N^{\text{new}}$ since we already know by Corollary 2 above that $\mathcal{N}(\mathbb{V}_N)$ generates $\mathbb{V}_N^{\text{new}}$.

Corollary 2.63 *If $0 \neq f \in \mathbb{V}_N$ is a $\mathbb{T}_N(ND)$-eigenform, then there exists a divisor $M|N$ and a normalized newform $g \in \mathcal{N}(\mathbb{V}_M)$ which has the same $\mathbb{T}_N(ND)$-character as f.*

Proof. Let $\chi : \mathbb{T}(ND) \to \mathbb{C}^\times$ denote the character defined by f, i.e. $f|T = \chi(f)f, \forall T \in \mathbb{T}(ND)$. Now since each term in the formula (2.50) for $\mathbb{V}^{\text{old}}$ is a $\mathbb{T}(ND)$-module, we see that χ has to appear in some $\beta_d^* \mathbb{V}_M^{\text{new}}$ for some $dM|N$, $M \neq N$, and hence also in $\mathbb{V}_M^{\text{new}} \simeq \beta_d^* \mathbb{V}_M^{\text{new}}$. Thus, there is a non-zero $\mathbb{T}(ND)$-eigenform $g \in \mathbb{V}_M^{\text{new}}$ with $g|T = \chi(T)g$, for all $T \in \mathbb{T}(ND)$. By Corollary 1, $g = ch$ is a scalar multiple of some $h \in \mathbb{V}_M^{\text{new}}$, and so the assertion follows.

Corollary 2.64 *The space $\mathbb{V}_N^{\text{new}}$ is the sum of the eigenspaces of $T_N(ND)$ whose eigencharacters occur with multiplicity one, whereas the space $\mathbb{V}_N^{\text{old}}$ is the sum of the eigenspaces whose eigencharacters appear with multiplicity greater than one.*

Proof. By Theorem 2.61, each $\mathbb{T}(ND)$-eigenspace of $\mathbb{V}_N^{\text{new}}$ is one-dimensional. On the other hand, if $\chi : \mathbb{T}(ND) \to \mathbb{C}$ is a character which appears in $\mathbb{V}_N^{\text{old}}$, i.e. there is an $f \in \mathbb{V}_N^{\text{old}}$ such that $f|T = \chi(T)f, \forall T \in \mathbb{T}(ND)$, then by Corollary 2 there is a $g \in \mathcal{N}(\mathbb{V}_M)$, $M|N$, $M \neq N$, with character χ, and then $g|\beta_{N/M}$ and g are two linearly independent forms in $\mathbb{V}_N^{\text{old}}$ with the same character χ.

Corollary 2.65 *Each $f \in \mathcal{N}(\mathbb{V}_N)$ is a $(\mathbb{T}_N \cup \mathbb{T}_N^*)$-eigenform, where $\mathbb{T}_N^* = \{T^* \in End(\mathbb{V}_N) : T \in \mathbb{T}_N\}$ is the subalgebra consisting of all adjoints of $\mathbb{T}_N$. Thus we have*

$$f|_k T_n^* = \overline{a_n(f)} f, \quad \text{for all } n \geq 1. \tag{2.51}$$

Moreover, if $f^* := \sum_{n\geq 1} \overline{a_n(f)} q^n$, then $f^* \in \mathcal{N}(\mathbb{V}_N)$ and we have

$$f|_k w_N = c f^*, \quad \text{for some } c \in \mathbb{C}^\times. \tag{2.52}$$

In particular, w_N maps $\mathbb{V}_N^{\text{new}}$ (and $\mathbb{V}_N^{\text{old}}$) into itself.

Proof. Since $\mathbb{T}_N(N)$ is $*$-closed, we see that $\mathbb{T}_N(N) \subset \mathbb{T}_N \cap \mathbb{T}_N^*$. Now since $\mathbb{T}_N$ is commutative, so is $\mathbb{T}_N^*$, and hence T^* commutes with $\mathbb{T}_N(N)$, if $T \in \mathbb{T}_N$. Thus, $f|T^*$ is in the $\mathbb{T}_N(N)$-eigenspace of f, and so by the Multiplicity-One Theorem we have $f|T^* = c_T f$, for some $c_T \in \mathbb{C}$. This proves the first statement.

Thus, $f|T_n^* = c_n f$, for some $c_n \in \mathbb{C}$. Since also $f|T_n = a_n(f)f$, we obtain $c_n\langle f, f\rangle = \langle f|T_n^*, f\rangle = \langle f, f|T_n\rangle = \langle f, a_n(f)f\rangle = \overline{a_n(f)}\langle f, f\rangle$, and so $c_n = \overline{a_n(f)}$. Thus, (2.46) holds.

To prove (2.47), we first recall that w_N normalizes $\Gamma_1(N)$; cf. Remark 2.46c). Furthermore, since $\beta_M w_N \beta_N^{-1} = w_M$, $\forall M|N$, we see that w_N maps $\mathbb{V}_{N^{\text{old}}}$ into itself, and hence also $\mathbb{V}_N^{\text{new}} = (\mathbb{V}_N^{\text{old}})^\perp$, because $w_N^* = -w_N$. Thus $g := f|_k w_N \in \mathbb{V}_N^{\text{new}}$. Furthermore, since $T_n^* = w_N T_n w_N^{-1}$ by (2.36), we obtain $g|_k T_n = f|_k w_N T_n = f|T_n^* w_N = \overline{a_n(f)} f|_k w_N = \overline{a_n(f)} g$. This means that g is a $\mathbb{T}$-eigenform with T_n-eigenvalue $\overline{a_n(f)}$, and so (2.52) holds with $c = a_1(g)$; cf. Proposition 2.36. Thus $f^* \in \mathbb{V}_N^{\text{new}}$ is a $\mathbb{T}$-eigenform, and hence $f^* \in \mathcal{N}(\mathbb{V}_N)$.

Remark. *a) Note that while the algebra $\langle \mathbb{T}_N, \mathbb{T}_N^*\rangle$ generated by $\mathbb{T}_N$ and $\mathbb{T}_N^*$ is $*$-closed, it is not commutative in general, for otherwise $\mathbb{V}_N$ would have a basis consisting of $\mathbb{T}$-eigenforms.*

b) It is immediate from Corollaries 1 and 4 that $\mathbb{V}_N^{\text{new}}$ and $\mathbb{V}_N^{\text{old}}$ are $\mathbb{T}_N$-modules as well as $\mathbb{T}_N^*$-modules.

Notation. For any $M|N$ and $f \in \mathcal{N}(\mathbb{V}_M)$, let

$$S_f(N) \;=\; S_f(\mathbb{V}_N) \;=\; \sum_{d|N/M} \mathbb{C} f|_k \beta_d \;=\; \bigoplus_{d|N/M} \mathbb{C} f|_k \beta_d;$$

clearly $\dim S_f(N) = \sigma_0(\frac{N}{M})$ = number of divisors of $\frac{N}{M}$. Furthermore, we let $\mathcal{N}^*(\mathbb{V}_N) := \bigcup_{M|N} \mathcal{N}(\mathbb{V}_M)$ denote the set of normalized newforms of all levels $M|N$.

It is immediate from equation (2.39) that every $g \in S_f(N)$ has the same $\mathbb{T}(N)$-eigenvalues as f, and hence is a subspace of the $\mathbb{T}(N)$-eigenspace defined by f. However, the fact that $S_f(N)$ is actually the full $\mathbb{T}(N)$-eigenspace seems to lie deeper, and requires a further result, whose proof will be sketched in the next section.

Theorem 2.66 *Suppose $f \in \mathbb{V}_N^{\text{new}}$ a $\mathbb{T}_N$-eigenform and $g \in \mathbb{V}_M$ is a $(\mathbb{T}_M \cup \mathbb{T}_M^*)$-eigenform such that $a_n(f) = a_n(g)$, for all $(n, D) = 1$ (for some $D \geq 1$), then $f = g$ and $N = M$.*

Corollary 2.67 *Let $D \geq 1$. For any $f \in \mathcal{N}^*(\mathbb{V}_N)$, the space $S_f(N)$ is the $\mathbb{T}_N(ND)$-eigenspace defined by f and hence*

$$\mathbb{V}_N = \bigoplus_{M|N} \bigoplus_{f \in \mathcal{N}(\mathbb{V}_M)} S_f(N) = \bigoplus_{f \in \mathcal{N}^*(\mathbb{V}_N)} S_f(N). \tag{2.53}$$

is the decomposition of $\mathbb{V}_N$ into distinct $\mathbb{T}(ND)$-eigenspaces. Furthermore, each $S_f(N)$ is a $\mathbb{T}_N$-module and a $\mathbb{T}_N^*$-module.

Proof. First note that the decomposition (2.50), together with Corollary 1 of Theorem 2.61 (applied to all levels $M|N$), shows that $\mathbb{V}_N = \sum_{f \in \mathcal{N}^*(\mathbb{V}_N)} S_f(N)$.

Next we observe that by (a slight generalization of) Proposition 2.45, $\mathbb{V} := \mathbb{V}_N$ has a basis consisting of $\mathbb{T}(ND)$-eigenforms, or, equivalently, $\mathbb{V}$ has a (unique) decomposition into $\mathbb{T}(ND)$-eigenspaces $\mathbb{V}_\chi := \{g \in \mathbb{V} : g|T = \chi(T)g, \forall T \in \mathbb{T}(ND)\}$, i.e. $\mathbb{V} = \bigoplus_\chi \mathbb{V}_\chi$, where $\chi : \mathbb{T}(ND) \to \mathbb{C}$ runs over all characters of $\mathbb{T}(ND)$.

In addition, we note that for each $f \in \mathcal{N}^*(\mathbb{V}_N)$ we have by (2.39) that $S_f \subset \mathbb{V}_{\chi_f}$, where $\chi_f : \mathbb{T}(ND) \to \mathbb{C}$ denotes the character defined by f i.e. by $f|T = \chi_f(T)f$, $\forall T \in \mathbb{T}(ND)$. Thus we have

$$\mathbb{V} = \sum_{f \in \mathcal{N}^*(\mathbb{V}_N)} S_f(N) \subset \sum_{f \in \mathcal{N}^*(\mathbb{V}_N)} \mathbb{V}_{\chi_f} \subset \bigoplus_\chi \mathbb{V}_\chi = \mathbb{V},$$

and so all inclusions have to be equalities. In particular, for each character χ we have $\mathbb{V}_\chi = \sum_{\chi_f = \chi} S_f(N)$, where the sum is over all $f \in \mathcal{N}^*(\mathbb{V}_N)$ such that $\chi_f = \chi$.

However, the characters χ_f are all pairwise distinct because if $\chi_f = \chi_g$, where $f, g \in \mathcal{N}^*(\mathbb{V}_N)$, then by Corollary 4 of Theorem 2.61, the hypotheses of Theorem 2.66 are satisfied, and so $f = g$, as claimed. Thus, every character χ is of the form $\chi = \chi_f$, for a unique $f \in \mathcal{N}^*(\mathbb{V}_N)$, and each $S_f(N) = \mathbb{V}_{\chi_f}$ is a complete $\mathbb{T}(ND)$-eigenspace. Furthermore, the decomposition (2.53) holds.

Finally, to see that $S_f(N)$ is a $\mathbb{T}$-module, let $T \in \mathbb{T}$. Since T commutes with $\mathbb{T}(ND)$, we have that $f|T \in \mathbb{V}_{\chi_f} = S_f(N)$, and so $S_f(N)$ is a $\mathbb{T}$-module. Similarly, $S_f(N)$ is also a $\mathbb{T}^*$-module, as a variant of the proof of Corollary 4 above shows.

Corollary 2.68 *Suppose that $g \in \mathbb{V}_N$ is a $\mathbb{T}(ND)$-eigenform for some $D \geq 1$. Then there exists a unique divisor $M|N$ and a unique normalized newform $f \in \mathcal{N}(\mathbb{V}_M)$ of level M such that f and g have the same $\mathbb{T}(ND)$-eigenvalues.*

Proof. Since (2.53) is the decomposition of $\mathbb{V}_N$ into distinct $\mathbb{T}(ND)$-eigenspaces, we have that $g \in S_f(N)$, for a unique $M|N$ and a unique $f \in \mathcal{N}(\mathbb{V}_M)$.

Corollary 2.69 *Let $f \in \mathbb{V}_N$. Then f is a normalized newform (of level N) if and only if f is a $\mathbb{T}_N \cup \mathbb{T}_N^*$-eigenform.*

Proof. If $\mathcal{N}(\mathbb{V}_N)$, then f is a $(\mathbb{T}_N \cup \mathbb{T}_N^*)$-eigenform by Corollary 4 of Theorem 2.61. Conversely, if f is a $(\mathbb{T}_N \cup \mathbb{T}_N^*)$-eigenform, then by Corollary 2 we have $f \in S_g(N)$, for some $g \in \mathcal{N}(\mathbb{V}_M)$, with $M|N$. By Theorem 2.66 it then follows that $f = g$ and $M = N$, so $f \in \mathcal{N}(\mathbb{V}_N)$.

Corollary 2.70 *Let $f \in \mathbb{V}_N$. Then f is a normalized newform (of level N) if and only if f and $f|w_N$ are both $\mathbb{T}_N$-eigenforms.*

Proof. Since f is a $\mathbb{T}^*$-eigenform if and only if $f|w_N$ is a $\mathbb{T}$-eigenform (cf. proof of Corollary 4 of Theorem 2.61), Corollary 4 is just a restatement of Corollary 3.

Remark. *Note that the decomposition (2.53) shows that the set*

$$\mathcal{B}(\mathbb{V}_N) := \bigcup_{M|N} \bigcup_{f \in \mathcal{N}(\mathbb{V}_M)} \{f|_k \beta_d\}_{d|\frac{N}{M}}$$

is a basis *of $\mathbb{V}_N$. Thus, $\mathbb{V}_N$ has a "canonical basis" consisting of normalized newforms of all levels, together with certain "twists" of these with respect to the operators β_d, $d|N$.*

2.4.4 Sketch of Proofs

We now sketch the main ideas of the proofs of Theorems 2.58 and 2.66, following (in part) the presentation given in Lang[La], pp. 126–137 and Miyake[Mi], pp. 153–175.

Step 1. $\mathbb{V}(DN) = \mathbb{V}(N)$, for all $D \geq 1$.

The proof of this step uses the following results.

Lemma 2.71 (Hecke) *If $(d, N) = 1$, then $M_k(\Gamma_1(N)) \cap M_k(\Gamma_1(N))|_k \beta_d = \{0\}$.*

Proof. This is a special case of Miyake[Mi], Lemma 4.6.3; cf. also Lang[La], proof of Theorem 4.1.

Lemma 2.72 *If f is a homomorphic function on $\mathfrak{H}$ such that $f(z+1) = f(z)$ and such that $f|_k \beta_d \in M_k(\Gamma_1(N))$, for some $d \geq 1$, then $f \in M_k(\Gamma_1(N))$*

Proof. [Mi], first part of proof of Theorem 4.6.4.

Lemma 2.73 *If $f = \sum a_n q^n \in M_k(N, \chi)$, then $f_{(D)} := \sum_{(n,D)=1} a_n q^n \in M_k(ND^2, \chi)$, for any $D \geq 1$.*

Proof. [Mi], Lemma 4.6.5.

Remark. *a) If $f \in \mathbb{V}$, then by definition $f \in \mathbb{V}(D) \Leftrightarrow f_{(N)} = 0$.*

b) For any prime p we have $f - f_{(p)} = \sum_{p|n} a_n(f) q^n = h|_k \beta_p$, where h is a suitable power series in q.

Proof of Step 1. By definition, $\mathbb{V}(N) \subset \mathbb{V}(ND)$. Suppose that the inclusion is proper, i.e. that there exists $f \in \mathbb{V}(ND) \setminus \mathbb{V}(N)$. Then $f_{(ND)} = 0$ but $f_{(N)} \neq 0$, so there exists a divisor $D_1|D$ and a prime $p \nmid ND_1$ such that $g := f_{(ND_1)} \neq 0$ but $g_{(p)} = f_{(ND_1p)} = 0$. Thus $g = g - g_{(p)} = h|\beta_p$, for some power series h in q. Now $g \in S_k(\Gamma_1(ND_1^2))$ by Lemma 2.73, so also $h \in M_k(\Gamma_1(ND_1^2))$ by Lemma 2.72. But then $g \in M_k(\Gamma_1(ND_1^2)) cap M_k(\Gamma_1(ND_1^2))|\beta_p = \{0\}$ by by Lemma 2.71 (since $p \nmid ND_1^2$), contradiction. Thus, no such f exists, i.e. $\mathbb{V}(ND) = \mathbb{V}(N)$.

Step 2. For all primes $p|D$ with $p^2|N$ we have $\mathbb{V}_N(D) \subset \mathbb{V}_N(D/p) + \mathbb{V}_{N/p}|\beta_p$.

For this, we shall use:

Lemma 2.74 *a) If $p|N$ and $f \in M_k(\Gamma_1(N)))$, then $f_{(p)} = f - p^{-k/2} f|_k T_p \beta_p$. Thus $f_(p) \in M_k(\Gamma_1(Np))$.*

b) If $p^2|N$ and $f \in M_k(\Gamma_1(N)))$, then $f|T_p \in M_k(\Gamma_1(N/p))$ and hence $f_{(p)} \in M_k(N)$.

Proof. a) Recall that $f|_k\beta_p(z) = p^{k/2} f(dz)$. Moreover, since $p|N$, we have $T_p = U_p$ (cf. Proposition 2.34). Thus, $a_n(f|_k T_p \beta_p) = 0$, if $p \nmid n$ and $a_n(f_k T_p \beta_p) = p^{k/2} a_n(f)$, if $p|n$, and so the assertions follow.

b) [La], Lemma 6 (p. 133).

Proof of step 2. Let $f \in \mathbb{V}(D)$. Then $g := f_{(p)} \in \mathbb{V}_N$ by Lemma 2.74b). Moreover, since $g_{(D/p)} = f_{(D)} = 0$, we see that $g \in \mathbb{V}_N(D/p)$. On the other hand, by Lemma 2.74 we have $f = g + h|\beta_p$ with $h = p^{1-k/2} f|T_p \in M_k(\Gamma_1(N/p))$, and so step 2 follows.

Step 3. For all squarefree $D|N$ and primes $p|D$ we have $\mathbb{V}_N(D) \subset \mathbb{V}_N(D/p) + \mathbb{V}_{N/p}|\beta_p$.

Lemma 2.75 *Let $p|N$ and $f \in M_k(N, \chi)$, where χ is not a character mod N/p. If $f = g|\beta_p$, for some g, then $f = 0$.*

Proof. [La], Lemma 4 (p. 131).

Lemma 2.76 *Suppose $p|N$, and χ is a character mod N/p. Then the operator $\tilde{T}_p^N := T_{\Gamma_0(N),\Gamma_0(N/p)}(\alpha_p)$ defines a linear map*

$$\tilde{T}_p^N : M_k(N, \chi) \to M_k(N/p, \chi)$$

with the following properties:

$$f|_k \tilde{T}_p^{ND} = f|_k \tilde{T}_p^N, \quad \text{if } f \in M_k(N, \chi),\ p \nmid D. \tag{2.54}$$

$$f|\beta_q \tilde{T}_p^{Nq} = f|k\tilde{T}_p^N \beta_q, \quad \text{if } q \neq p \tag{2.55}$$

$$f|\beta_p \tilde{T}_p^N = p^{k/2} c_p f, \quad \text{if } f \in M_k(N/p, \chi) \tag{2.56}$$

with $c_p = 1$ if $p^2|N$ and $c_p = 1 + \frac{1}{p}$ otherwise.

Proof. The properties (2.54) and (2.55) are proved in [Mi], Lemma 4.6.6, and property (2.56) is proved on the top of p. 161 of [Mi].

Lemma 2.77 *For any $D \geq 1$ we have*

$$\mathbb{V}_N(D) = \bigoplus_{\chi} \mathbb{V}_{N,\chi}(D),$$

where the sum is over all Dirichlet characters mod N and $\mathbb{V}_{N,\chi}(D) = \mathbb{V}_N(D) \cap M_k(N,\chi)$.

Proof. By step 1, we may assume that $D|N$. Then $f_{(D)}|_k\sigma_a = (f|\sigma_a)_{(D)}$ (cf. [La], Lemma 3, p. 131), and so $\mathbb{V}(D)$ is stable under the action of the σ_a's and so the assertion follows.

Proof of step 3. Let $f \in \mathbb{V}(D)$; by Lemma 2.77 we may assume that $f \in \mathbb{V}_{N,\chi}(D)$, for some Dirichlet character χ. Put $D_1 = \frac{N}{p}$ and $g = f_{(D_1)}$, $h = f - g$. Then $g, h \in M_k(N_1, \chi)$, where $N_1 = ND_1^2$. Since $g_{(p)} = f_{(D)} = 0$, we see that $g = g_p|\beta_p$, where g_p is some power series in q. If χ is not a character modulo N_1/p, then $g_p = 0$ and then $f_{(D_1)} = g = 0$, i.e. $f \in \mathbb{V}(D_1)$, and we are done.

Thus, assume that χ is defined mod N_1/p (and hence also modulo N/p). Then by ? $g_p \in M_k(N_1/p, \chi)$.

Put $f_p := f|_k\tilde{T}_p^N$; we claim that $f - f_p|\beta_p \in \mathbb{V}_N(D_1)$.

2.4.5 Exercises

1. Prove that the Hecke algebra $\mathbb{T} \subset \mathrm{End}_{\mathbb{C}}(S_k(\Gamma)))$ is semi-simple if and only if $S_k(\Gamma)$ has a basis consisting of $\mathbb{T}$-eigenforms.

2. (a) Let p be a prime and let $\mathbb{T}^{(p)} \subset \mathrm{End}_{\mathbb{C}}(S_2(\Gamma_0(p^2)))$ be the Hecke algebra (over $\mathbb{C}$) generated by the Hecke operators T_n with $p \not| n$. Show that

$$\dim \mathbb{T}^{(p)} = g(X_0(p^2)) - g(X_0(p)).$$

   (b) Generalize part (a) to $S_2(\Gamma_0(p^r))$.

Bibliography

[Ah] L. Ahlfors, *Complex Analysis*. Addison-Wesley, Reading, 1965.

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

[Bi] F. Bien, Construction of telephone networks by group representations. *Notices AMS* **36**(1989), 5–22.

[Bo] A. Borel, S. Chowla, C.S. Herz, K. Iwasawa, J.-P. Serre, *Seminar on Complex Multiplication*. Springer Lecture Notes **21** (1966).

[BCDT] C. Breuil, B. Conrad, F. Diamond, R. Taylor, On the modularity of elliptic curves over $\mathbb{Q}$: with 3-adic exercises. *J. Am. Math. Soc.* **14** (2001), 843–939.

[CN] J.H. Conway, S.P. Norton, Monstrous Moonshine. *Bull. London Math. Soc.* **11** (1979), 308–339.

[CS] J.H. Conway, N.J.A. Sloane, *Sphere Packings, Lattices and Groups.* Springer-Verlag, New York, 1988.

[De1] P. Deligne, Formes modulaires et représentations ℓ-adiques. *Sem. Bourbaki* 1968/69, exp. 355; Springer Lecture Notes **179** (1971).

[De2] P. Deligne, La Conjecture de Weil I, II. *Publ. IHES* **43** (1974), 273-307; **52** (1980), 137 -152.

[Dij] R. Dijkgraaf, Mirrow symmetry and elliptic curves. In: *The Moduli Space of Curves* (R. Dijkgraaf, C. Faber, G. van der Geer, eds.) Birkhäuser, Boston, 1995, pp. 149–163.

[Fo] O. Forster, *Lectures on Riemann Surfaces.* Springer-Verlag, New York, 1982.

[Gl] J.W.L. Glaisher, On the square of the series in which the coefficients are the sums of the divisors of the exponents. *Messenger of Math.* **14** (1884/85), 156–163.

[Gu] R.C. Gunning, *Lectures on Modular Forms*. Princeton Univ. Press, Princeton, 1962.

[HW] G.H. Hardy, E.M. Wright, *An Introduction to the Theory of Numbers.* (4th ed.). Oxford U. Press, London, 1968.

[He] E. Hecke, Über Modulfunktionen und die Dirichletschen Reihen mit Eulerscher Produktentwicklung. I. II. *Math. Ann.* **114** (1937), 1–28, 316–357 = Math. Werke, pp. 644–703.

[Hua] Hua Loo Keng, *Introduction to Number Theory.* Springer-Verlag, Berlin, 1982.

[IR] K. Ireland. M. Rosen, *A Classical Introduction to Modern Number Theory.* Springer-Verlag, New York, 1982.

[Iw] H. Iwaniec, *Topics in Classical Automorphic Forms.* Amer. Math. Soc., Providence, 1997.

[Kl] F. Klein, Zur [Systematik der] Theorie der Modulfunktionen. *Sitzber. Akad. Wiss. München*, 1879 = Gesammelte Math. Abhandlungen III, Springer, Berlin, 1923, pp. 168–178.

[Kn] M. Knopp, *Modular functions in Analytic Number Theory.* Markham Publ. Co., Chicago, 1970.

[Ko] N. Koblitz, *Introduction to Elliptic Curves and Modular Forms.* Springer-Verlag, New York, 1984.

[KZ] M. Kaneko, D. Zagier, A generalized Jacobi theta function and quasimodular forms. In: *The Moduli Space of Curves* (R. Dijkgraaf, C. Faber, G. van der Geer, eds.) Birkhäuser, Boston, 1995, pp. 165–172.

[Land] P.S. Landweber (ed.), *Elliptic Curves and Modular Forms in Algebraic Topology (Proceedings, Princeton, 1986).* Springer Lecture Notes **1326** (1988).

[La0] S. Lang, *Elliptic Functions.* Addison-Wesley, Reading, MA, 1973.

[La] S. Lang, *Introduction to Modular Forms.* Springer-Verlag, Berlin, 1976.

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

[Mc] I.G. Macdonald, Affine root systems and Dedekind's η-function. *Invent. Math.* **15** (1972), 91–143.

[Mi1] Miyake, On automorphic forms on GL_2 and Hecke operators. *Ann. Math.* **94** (1971), 174-189.

[Mi] T. Miyake, *Modular Forms.* Springer-Verlag, Berlin, 1989.

[Ne] M. Newman, *Integral Matrices.* Acandemic Press, New York, 1972.

[Pe] H. Petersson, Konstruktion der sämtlichen Lösungen einer Riemannschen Funktionalgleichung durch Dirichlet-Reihen mit Eulerscher Produktentwicklung. I. *Math. Ann.* **116** (1930), 401–412.

[Ra] S. Ramanujan, On certain arithmetical functions. *Trans. Cambridge phil. Soc.* **22** (1916), 159–184 = *Collected Papers*, No. **18**, 136–162.

[Sa] P. Sarnak, *Some Applications of Modular Forms.* Cambridge University Press, Cambridge, 1990.

[Sch] B. Schoeneberg, *Elliptic Modular Functions.* Springer-Verlag, Berlin, 1974.

[Se1] J.-P. Serre, *A Course in Arithmetic.* Springer-Verlag, New York, 1973.

[Se2] J.-P. Serre, Quelques applications du théorème de densité de Chebotarev. *Publ. Math. IHES* **54** (1981), 123–201 = Œuvres/Collected Papers III, Springer-Verlag, Berlin, 1986, pp. 563–641.

[Sh] G. Shimura, *Introduction to the Arithmetic Theory of Automorphic Functions.* Iwanami Shoten, 1971.

[Si] C.L. Siegel, *Topics in Complex Function Theory I.* Wiley, New York, 1969.

[ST] J. Silverman, J. Tate, *Rational Points on Elliptic Curves.* Springer-Verlag, New York, 1992.

[Si1] J. Silverman, *The Arithmetic of Elliptic Curves.* Springer-Verlag, New York, 1986.

[Si2] J. Silverman, *Advanced Topics in the Arithmetic of Elliptic Curves.* Springer-Verlag, New York, 1994.

[Sp] G. Springer, *Introduction to Riemann Surfaces.* Addison-Wesley, Reading, MA, 1957.

[SwD] H.P.F Swinnerton-Dyer, On ℓ-adic representations and congruences for coefficients of modular forms. In: *Modular Functions of One Variable III*, Springer Lecture Notes **350** (1973), pp. 1–55.

[We] H. Weber, *Lehrbuch der Algebra III.* 2nd ed. 1908. Reprint: Chelsea, New York.

[We1] A. Weil, Jacobi sums as "Grössencharacters". *Trans. AMS* **73**, 487–495 = Œuvres Scientitifiques/Collected Papers II, Springer-Verlag, 1979, pp. 63–71.

[We2] A. Weil, Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen. *Math. Ann.* **168**, 149–156 = Œuvres Scientitifiques/Collected Papers III, Springer-Verlag, 1979, pp. 165–172.

[Wi] A. Wiles, Modular elliptic curves and Fermat's Last Theorem. *Ann. Math.* **141** (1995), 443-551.