1

The Basic Modular Forms of the Nineteenth Century

1.1 The Modular Group

Nineteenth-century mathematicians did not always give precise definitions in their work, though from the context and examples, their meaning is usually quite discernible. We here offer some examples of the basic concepts created and studied by these mathematicians, such as modular forms and the multiplier system, and discuss how these ideas would now be classified and defined.

The modular group $SL_2(\mathbb{Z})$ consists of all 2×2 matrices with integer entries and determinant 1:

$$SL_2(\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, a, b, c, d \in \mathbb{Z},\ ad - bc = 1 \right\}. \tag{1.1}$$

It can be deduced from results proved in 1775 by Lagrange,[1] in connection with the reduction of quadratic forms, that the modular group can generated by

$$S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{1.2}$$

Here the notation for S and T follows early researchers such as Mordell and Rademacher; some recent works have interchanged S and T.

Observe that S and T generate $SL_2(\mathbb{Z})$ by an application of the Euclidean algorithm. Given a 2×2 matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, suppose $c \neq 0$ and $|a| \geq |c|$. In that case, there exists an integer n such that

$$S^n \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a + cn & b + dn \\ c & d \end{pmatrix} \equiv \begin{pmatrix} a_1 & b_1 \\ c & d \end{pmatrix}, \tag{1.3}$$

where $0 \leq a_1 < |c|$. By multiplying the matrix on the far right-hand side of (1.3) by T, we can move the entries c and d to the first row:

$$T \begin{pmatrix} a_1 & b_1 \\ c & d \end{pmatrix} = \begin{pmatrix} -c & -d \\ a_1 & b_1 \end{pmatrix}. \tag{1.4}$$

[1] See Weil (1983) pp. 216–217 and 321.

Now if $a_1 \neq 0$, then the calculation indicated in (1.3) and (1.4) can be repeated until we obtain a matrix in which the entry in the first column and second row is 0. Since the determinant of all matrices in the modular group is 1, the matrix containing this entry must take the form

$$\begin{pmatrix} \pm 1 & \pm k \\ 0 & \pm 1 \end{pmatrix}.$$

Next, note that

$$S^{-k}\begin{pmatrix} \pm 1 & \pm k \\ 0 & \pm 1 \end{pmatrix} = \begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}$$

and

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

proving that S and T as given in (1.2) are the generators of $SL_2(\mathbb{Z})$.

Let $\mathbb{H}$ denote the set of all complex numbers $x + iy$, where x is a real number and $y > 0$. The set $\mathbb{H}$ is called the upper half-plane. The elements of $SL_2(\mathbb{Z})$ act upon the elements of $\mathbb{C} \cup \{\infty\}$, on the complex plane, and on the point at infinity by means of the equation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z) = \frac{az+b}{cz+d}, \quad \text{where} \quad z \in \mathbb{C} \cup \{\infty\}. \tag{1.5}$$

Note that if $z \in \mathbb{H}$, then

$$\operatorname{Im}\frac{az+b}{cz+d} = \frac{\operatorname{Im} z}{|cz+d|^2} > 0. \tag{1.6}$$

Thus, (1.5) serves also to define the action of $SL_2(\mathbb{Z})$ on $\mathbb{H}$.

The modular group and some of its subgroups were used in unpublished works of Gauss in his theory of binary quadratic forms and in his work on the arithmetic-geometric mean of two complex numbers. However, it seems that Dedekind was the first mathematician to explicitly define the modular group and its fundamental domain. He did this in a paper of 1877, and two years later, Felix Klein singled out the congruence subgroups of the modular group as the concept capable of bringing order and focus to the theory of modular functions.

Let $\Gamma = \Gamma(1)$ designate the modular group $SL_2(\mathbb{Z})$ and, for a positive integer N, let

$$\Gamma(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \;\middle|\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod N \right\}. \tag{1.7}$$

$\Gamma(N)$ is called the principal congruence subgroup of level N. Any subgroup of Γ containing $\Gamma(N)$ for some N is called a congruence subgroup of level N, or simply a congruence subgroup. Adolf Hurwitz, in his 1881 doctoral dissertation, explicitly defined the subgroup

$$\Gamma^0(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \;\middle|\; b \equiv 0 \bmod N \right\}. \tag{1.8}$$

Hurwitz also proved that the index of $\Gamma(N)$ in Γ was given by

$$N^3 \prod_{p|N} \left(1 - \frac{1}{p^2}\right), \tag{1.9}$$

and that the index of $\Gamma^0(N)$ in Γ was

$$N \prod_{p|N} \left(1 + \frac{1}{p}\right). \tag{1.10}$$

We observe that (1.10) also expresses the index of the subgroup

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \;\middle|\; c \equiv 0 \bmod N \right\}. \tag{1.11}$$

The subgroup (1.11) is called the Hecke subgroup of the modular group, even though it was used by Hurwitz, because Hecke employed it extensively in his researches. Hurwitz considered another useful congruence subgroup:

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \bmod N \right\}. \tag{1.12}$$

The $*$ notation indicates that the entry is arbitrary and does not satisfy any congruence relation mod N. Note that the index of $\Gamma_1(N)$ in Γ is

$$N^2 \prod_{p|N} \left(1 - \frac{1}{p^2}\right). \tag{1.13}$$

Let Γ' be a congruence subgroup of $\Gamma = SL_2(\mathbb{Z})$. Two points $z_1, z_2 \in \mathbb{H}$ are called Γ'-equivalent if there exists a $\gamma \in \Gamma'$ such that $\gamma(z_1) = z_2$. A fundamental domain for Γ' is a closed region F in $\mathbb{H}$ such that every point in $\mathbb{H}$ is Γ'-equivalent to a point in F and no two points in the interior of F are Γ'-equivalent. Sometimes the interior of F is called the fundamental domain.

Dedekind found a fundamental domain for the modular group Γ; in his 1877 paper, he gave a description but not a diagram. A year later, Klein drew a picture of this, as well as pictures of fundamental domains of several congruence subgroups. The fundamental domain of the modular group as described by Dedekind and diagramed by Klein is represented in Figure 1.1, where the fundamental domain comprises the region including and enclosed by the two dark vertical lines and the arc of the circle between them. Dedekind and Klein discussed the subgroup $\Gamma(2)$; Klein's diagram of its fundamental domain is given in Figure 1.2, in which the fundamental domain is composed of the region including and enclosed by the two dark vertical lines and the two semicircles between them. Interestingly, Gauss had earlier drawn a diagram for a fundamental domain of $\Gamma(2)$ in a manuscript he did not publish.

An important subgroup of Γ, connected with a theta constant, given later in (1.40), is the group generated by S^2 and T. This is a congruence subgroup of level 2 and index 3. Denoted in Hecke's notation by $G(2)$, this subgroup can be shown to be a conjugate of $\Gamma_0(2)$:

$$G(2) = S^{-1} T \,\Gamma_0(2) T^{-1} S. \tag{1.14}$$

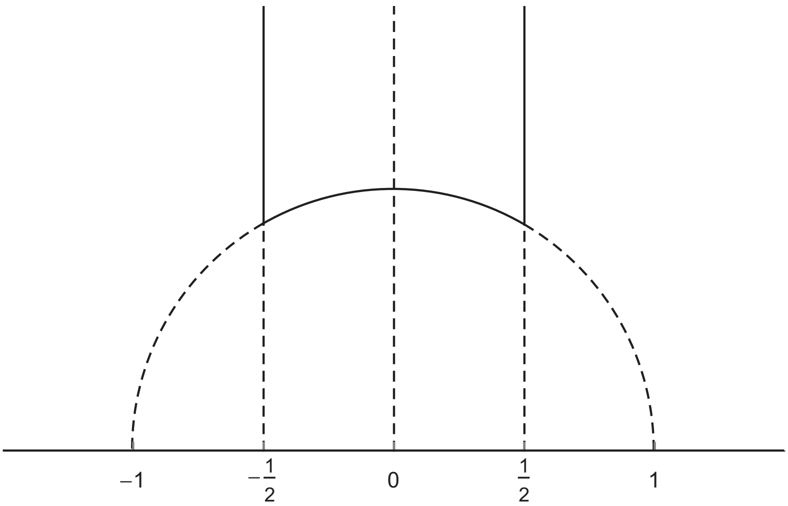

Figure 1.1. Fundamental Domain.

In their famous 1890–92 work on modular forms, Klein and R. Fricke denoted (1.14) by Γ_3, since it was a subgroup of Γ of index 3. A fundamental domain for $G(2)$, or Γ_3, is shown in Figure 1.3; this fundamental domain consists of the region including and enclosed by the two vertical lines at $+1$ and -1 and the semicircle between them.

The points corresponding to $\{\infty\} \cup Q$, where Q is the set of rational numbers, are called cusps. Observe that for any rational number $\frac{a}{c}$ in its reduced form, there exist integers b and d such that $ad - bc = 1$. Since

$$\frac{a\infty + b}{c\infty + d} = \frac{a}{c},$$

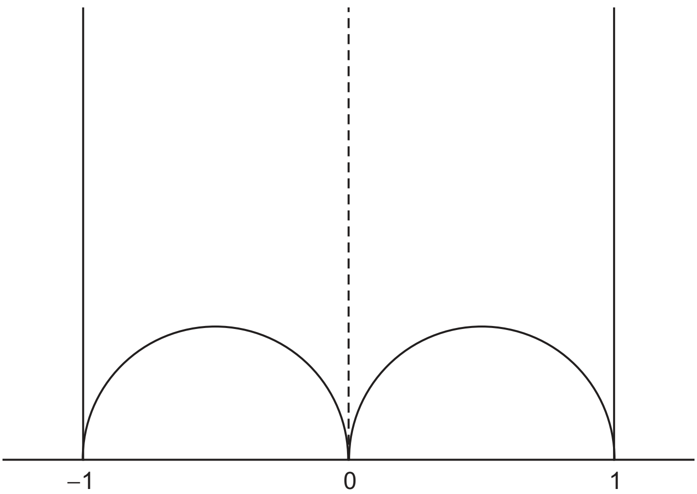

Figure 1.2. Fundamental Domain of $\Gamma(2)$.

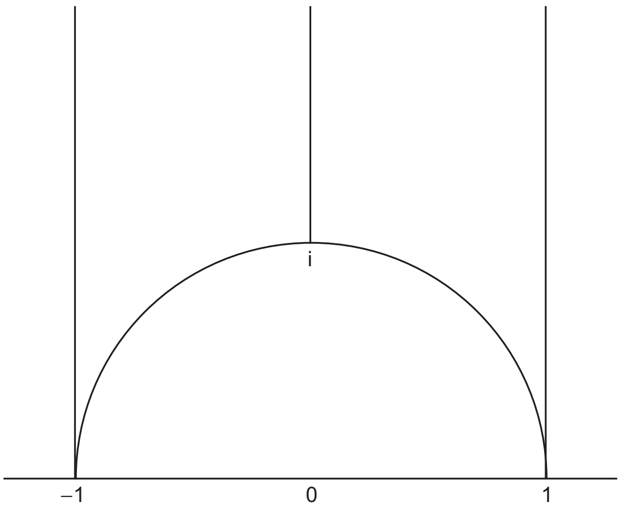

Figure 1.3. Fundamental Domain of $G(2)$.

it follows that a rational point is Γ-equivalent to infinity. Thus, the modular group permutes the cusps transitively. This may also be understood from the fact that the fundamental domain for Γ has only one cusp, that is, the cusp at infinity. A subgroup Γ' of Γ does not necessarily act transitively on the cusps. This means that for a given Γ', there is usually more than one Γ'-equivalence class among the cusps $\{\infty\} \cup Q$. These equivalence classes are called cusps of Γ'. Figure 1.2 illustrates that there are three inequivalent cusps of $\Gamma(2)$: $0, -1, \infty$. Observe that -1 is equivalent to 1, since $S^2 \in \Gamma(2)$ and $S^2(-1) = 1$. Note that $G(2)$ has two inequivalent cusps, one at ∞ and one at -1.

1.2 Modular Forms

Eisenstein, in his theory of elliptic functions, considered the series

$$G_k(\omega_1, \omega_2) = \sum_{m,n}{}' \frac{1}{(m\,\omega_1 + n\,\omega_2)^k}, \tag{1.15}$$

where $k \geq 2$ was an integer and ω_1, ω_2 were complex numbers such that $\text{Im}\, \frac{\omega_1}{\omega_2} > 0$. The summation was to be taken over all integers m, n except $m = n = 0$. He showed that the series converged absolutely when $k \geq 3$. This implied that, with integers a, b, c, d such that $ad - bc = 1$, if

$$\omega_1' = a\,\omega_1 + b\,\omega_2$$
$$\omega_2' = c\,\omega_1 + d\,\omega_2,$$

and if $k \geq 3$, then

$$G_k(\omega_1', \omega_2') = G_k(\omega_1, \omega_2). \tag{1.16}$$

Eisenstein noted that $G_k(\omega_1, \omega_2) \equiv 0$ when $k \geq 3$, with k odd. Now if we write $\omega = \frac{\omega_1}{\omega_2}$ and $G_k(\omega) \equiv G_k(\omega, 1)$, then

$$G_k(\omega) = \sum_{m,n}{}' \frac{1}{(m\,\omega + n)^k} \tag{1.17}$$

and 1.16 can be rewritten as

$$G_k(\omega') = (c\,\omega + d)^k G_k(\omega), \tag{1.18}$$

where

$$\omega' = \frac{a\,\omega + b}{c\,\omega + d}, \quad ad - bc = 1. \tag{1.19}$$

Since $G_{2k}(\omega + 1) = G_{2k}(\omega)$, $G_{2k}(\omega)$ has a Fourier series expansion with period one. This Fourier series was found by Hurwitz in his doctoral dissertation:

$$\begin{aligned} G_{2k}(\omega) &= 2\zeta(2k) + 2\frac{(2\pi i)^{2k}}{(2k-1)!}\sum_{n=1}^{\infty} \sigma_{2k-1}(n) e^{2\pi i n\omega} \\ &= 2\zeta(2k)\left(1 + \frac{(-1)^k\, 4k}{B_{2k}} \sum_{n=1}^{\infty} \sigma_{2k-1}(n)\, e^{2\pi i n\omega}\right), \end{aligned} \tag{1.20}$$

where $\sigma_{2k-1}(n) = \sum_{d|n} d^{2k-1}$ and where B_{2k} were Bernoulli numbers defined by

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} B_n \frac{x^n}{n!}.$$

Note that $B_1 = -\frac{1}{2}, B_{2m+1} = 0, m \geq 1$. The series $G_{2k}(\omega)$ is called an Eisenstein series. The discriminant function

$$\Delta(\omega) = \big(60G_4(\omega)\big)^3 - 27\big(140G_6(\omega)\big)^2 \tag{1.21}$$

played an important role in Weierstrass's theory of elliptic functions. Weierstrass employed his theory of sigma functions (equivalent to theta functions) to show that

$$\Delta(\omega) = (2\pi)^{12} e^{2\pi i\omega} \prod_{n=1}^{\infty} (1 - e^{2\pi i n\omega})^{24}. \tag{1.22}$$

Observe that (1.18) implies that

$$\Delta\left(\frac{a\,\omega + b}{c\,\omega + d}\right) = (c\,\omega + d)^{12}\Delta(\omega), \quad ad - bc = 1. \tag{1.23}$$

As a consequence of (1.18) and (1.22), we can see that

$$j(\omega) = \frac{1728\big(60G_4(\omega)\big)^3}{\Delta(\omega)} \tag{1.24}$$

satisfies, for integers a, b, c, d with $ad - bc = 1$, the relation

$$j\left(\frac{a\,\omega + b}{c\,\omega + d}\right) = j(\omega). \tag{1.25}$$

The function (1.24) was considered by Hermite in a 1858 paper on modular equations. In 1877, Dedekind defined $v(\omega) = \frac{1}{1728} j(\omega)$ by means of a conformal mapping of the fundamental domain of $\Gamma(1)$. Klein published a paper in 1878 in which he independently rediscovered this function, (1.24), through invariant theory. From that time onward, the function became known as Klein's J function, where $J(\omega) = v(\omega) = \frac{1}{1728} j(\omega)$. It also appears that Gauss was moving toward this function, in the context of his work in the theory of quadratic forms, though he did not explicitly define it. Hermite gave the Fourier expansion

$$j(\omega) = e^{-2\pi i\,\omega} - 744 + 196884 e^{2\pi i\,\omega} + \cdots . \tag{1.26}$$

Another function, k^2, arose naturally within the theory of elliptic functions when the elliptic function was regarded as an elliptic integral. Hermite gave the relation between k^2 and j:

$$j = 256 \frac{(k^4 - k^2 + 1)^3}{k^2(1 - k^2)^2}. \tag{1.27}$$

In a paper of 1828, Jacobi proved that k^2 satisfied an equation similar to (1.25), provided that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(2). \tag{1.28}$$

Next note that we may define $\sqrt[12]{\Delta(\omega)}$, a function studied by Kiepert, Hurwitz, and Klein, from (1.22) by using the equation

$$\sqrt[12]{\Delta(\omega)} = 2\pi\, e^{\frac{\pi i \omega}{6}} \prod_{n=1}^{\infty} (1 - e^{2\pi i n\,\omega})^2. \tag{1.29}$$

By writing $U = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma(1)$, (1.23) will imply that

$$\sqrt[12]{\Delta(U\,\omega)} = v(U)(c\,\omega + d)\; \sqrt[12]{\Delta(\omega)}, \tag{1.30}$$

where $v(U)$ denotes a suitable twelfth root of unity. Now it is clear from (1.29) that

$$v(S) = e^{\frac{\pi i}{6}}, \tag{1.31}$$

where S is as in (1.2), because

$$\sqrt[12]{\Delta(\omega + 1)} = e^{\frac{\pi i}{6}}\; \sqrt[12]{\Delta(\omega)}. \tag{1.32}$$

Similarly, with T as in (1.2),

$$v(T) = e^{\frac{3\pi i}{2}}; \tag{1.33}$$

this follows from (1.30), since

$$\sqrt[12]{\Delta\left(-\frac{1}{\omega}\right)} = v(T)\,\omega\,\sqrt[12]{\Delta(\omega)}. \tag{1.34}$$

Taking $\omega = i$ in (1.34) yields the value of $v(T)$ in (1.33). Note that if U and V belong to the modular group $\Gamma(1)$, then (1.30) implies that

$$v(UV) = v(U)v(V). \tag{1.35}$$

Now since $\Gamma(1)$ is generated by S and T, it is possible in principle to find the value of $v(U)$ for any $U \in T(1)$. In fact, in his doctoral thesis of 1881, Hurwitz proved the transformation formula:

For any integers a, b, c, d with $ad - bc = 1$,

$$\sqrt[12]{\Delta(a\,\omega_1 + b\,\omega_2, c\,\omega_1 + d\,\omega_2)} = e^{\left((1-c^2)(db+3d(c-1)+c+3)+c(d+a-3)\right)\frac{i\pi}{6}}\,\sqrt[12]{\Delta(\omega_1, \omega_2)}. \tag{1.36}$$

Four years earlier, Dedekind gave the transformation formula for

$$\eta(\omega) = e^{\frac{\pi i \omega}{12}} \prod_{n=1}^{\infty} (1 - e^{2\pi i n \omega}), \tag{1.37}$$

except that he stated his result in terms of Dedekind sums, to be defined later. A famous particular case of Dedekind's result is

$$\eta\left(-\frac{1}{\omega}\right) = \sqrt{\frac{\omega}{i}}\,\eta(\omega). \tag{1.38}$$

The function in (1.37) is now known as the Dedekind η function. The theta constants can be written in terms of η functions; for example,

$$\begin{aligned}\theta_3(\omega) &= \prod_{n=1}^{\infty} \left(1 - e^{i\pi i n \omega}\right)\left(1 + e^{(2n-1)\pi i \omega}\right)^2 \\ &= \frac{\eta^2\left(\frac{\omega+1}{2}\right)}{\eta^2(\omega+1)}.\end{aligned} \tag{1.39}$$

Note that a theta function is a function of two variables: x and $q = e^{\pi i \omega}$, as given in (3.189). When $x = 0$, we sometimes refer to the function as a theta constant. Some powers of theta constants and their ratios are types of modular forms or modular functions.

The infinite product for $\theta_3(\omega)$ can also be expressed as a series:

$$\theta_3(\omega) = \sum_{n=-\infty}^{\infty} e^{\pi i n^2 \omega} = 1 + 2\sum_{n=1}^{\infty} e^{\pi i n^2 \omega}. \tag{1.40}$$

By the year 1808, Gauss knew that the infinite product (1.39) equalled the series (1.40), but he did not publish his results along these lines. In 1828, Jacobi rediscovered this fact. Gauss applied Fourier series to prove that

$$\theta_3\left(-\frac{1}{\omega}\right) = \sqrt{\frac{\omega}{i}}\,\theta_3(\omega). \tag{1.41}$$

It is not known exactly when Gauss proved (1.41). However, on the basis of the position of this work in his notebooks and other factors, it appears likely that he did this work around 1808. In 1817, Cauchy published (1.41) in a paper on waves using Fourier transforms. In 1828, Jacobi proved (1.41) based on his theory of elliptic functions. Note that

$$\theta_3(\omega+2) = \theta_3(\omega). \tag{1.42}$$

Gauss and Jacobi defined other theta functions:

$$\theta_0(\omega) = \sum_{n=-\infty}^{\infty} (-1)^n\, e^{\pi i n^2 \omega}, \tag{1.43}$$

$$\theta_1(\omega) = \sum_{n=-\infty}^{\infty} (-1)^n\, e^{2\pi i \left(\frac{2n+1}{2}\right)^2 \omega}, \tag{1.44}$$

$$\theta_2(\omega) = \sum_{n=-\infty}^{\infty} e^{2\pi i \left(\frac{2n+1}{2}\right)^2 \omega}. \tag{1.45}$$

They showed that, with S defined by (1.2),

$$\begin{aligned}
\theta_0(S\omega) &= \theta_0\left(-\frac{1}{\omega}\right) = \sqrt{\frac{\omega}{i}}\,\theta_2(\omega),\\
\theta_2(S\omega) &= \theta_2\left(-\frac{1}{\omega}\right) = \sqrt{\frac{\omega}{i}}\,\theta_0(\omega),\\
\theta_1(S\omega) &= \theta_1\left(-\frac{1}{\omega}\right) = -\sqrt{i\,\omega}\,\theta_1(\omega).
\end{aligned} \tag{1.46}$$

In 1828, Jacobi proved a result that can be used to show that $k^2(\omega)$ is invariant with respect to $\Gamma(2)$:

$$k^2(\omega) = \frac{\theta_2^4(\omega)}{\theta_3^4(\omega)} = e^{i\pi\,\omega} \prod_{n=1}^{\infty} \left(\frac{1+e^{2n\pi i\,\omega}}{1+e^{(2n-1)\pi i\,\omega}}\right)^8. \tag{1.47}$$

Key definitions serve to categorize functions such as the θ, η, Δ, k^2, and so on: A function is denoted a meromorphic modular form of integer weight k, belonging to $\Gamma(1)$ or with respect to $\Gamma(1)$, if: f is meromorphic on the upper half-plane $\mathbb{H}$; and

$$f\left(\frac{a\,\omega+b}{c\,\omega+d}\right) = (c\,\omega+d)^k\, f(\omega), \tag{1.48}$$

for $\omega \in \mathbb{H}$ with $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma(1)$; and $f(\omega)$ has a Fourier expansion

$$f(\omega) = \sum_{n=-m}^{\infty} a_n e^{2\pi i n \omega}. \tag{1.49}$$

In the case $k = 0$, $f(\omega)$ is called a modular function belonging to $\Gamma(1)$. Note that the function $z = e^{2\pi i \omega}$ maps the vertical strip of the fundamental domain of $\Gamma(1)$ into the unit disc $|z| < 1$; also note that

$$f(\omega) = \sum_{n=-m}^{\infty} a_n z^n.$$

When $m > 0$, we say that f has a pole of order m at ∞ or $i\infty$; when $-m \geq 0$, we say that f is holomorphic at ∞. Now when f is holomorphic in the upper half-plane and holomorphic at ∞ (or $i\infty$), then f is called a modular form of weight k belonging to $\Gamma(1)$. And if f is a modular form of weight k and $a_0 = 0$, then f is known as a cusp form of weight k.

Denoting the vector space of modular forms of weight k by $M_k(\Gamma)$ and the space of cusp forms of weight k by $S_k(\Gamma)$, we see that $S_k(\Gamma)$ is a subspace of $M_k(\Gamma)$. Observe that $j(\omega)$ is a modular function and that

$$G_k(\omega) \in M_k(\Gamma),$$
$$\Delta(\omega) \in S_{12}(\Gamma),$$
$$G_{k-12}(\omega)\Delta(\omega) \in S_k(\Gamma) \quad \text{for } k > 12.$$

Examples of functions in (1.36) and (1.38) lead us to an extension of modular forms. We denote v as a multiplier system of weight k for $\Gamma(1)$ if

$$\text{for } i = 1, 2, 3, \quad U_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}, \text{ and } U_3 = U_1 U_2,$$

with U_1, U_2 arbitrary elements of $\Gamma(1)$, we have $|v(U_i)| = 1$ and

$$v(U_3)(c_3\,\omega + d_3)^k = v(U_1)v(U_2)(c_1 U_2\,\omega + d_1)^k (c_2\,\omega + d_2)^k. \tag{1.50}$$

In the case when k is an integer, (1.50) reduces to

$$v(U_3) = v(U_1)v(U_2). \tag{1.51}$$

Next, let $v(S) = e^{2\pi i t}$, where t is a real number and S is given by (1.2). A function f, meromorphic on $\mathbb{H}$ and with the series expansion

$$f(\omega) = \sum_{n=-m}^{\infty} a_n e^{2\pi i (n+t)\omega}, \tag{1.52}$$

is designated a meromorphic modular form of weight k with multiplier system v if

$$f(U\,\omega) = v(U)(c\,\omega + d)^k f(\omega), \tag{1.53}$$