Modular Forms

Balázs Szendrői

September 1994

1 Introduction

The essay is concerned with modular forms of a complex variable **z** for the full modular group Γ, and also for some of its subgroups. The preliminary Chapter 2 is based on Gunning [1] and Serre [6]; Chapter 3 presents elementary results also based on Serre [6]. Chapter 4 shows the connection with elliptic curves, whereas in Chapter 5 we use the theory of Riemann-surfaces to get more general results, based on Gunning [1]. Finally, in Chapter 6 we present applications of the modular forms.

2 The modular group and its subgroups

2.1 The modular group Γ and its fundamental domain

Let $\mathbf{SL}_2(\mathbb{C})$ denote the set of 2×2 matrices, with complex coefficients, of determinant 1. This group acts on the Riemann sphere $\tilde{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ in the usual way: $g(z) = \dfrac{az+b}{cz+d}$ for $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}_2(\mathbb{C})$, $\mathbf{z} \in \mathbb{C}$; $g(\infty) = \dfrac{a}{c}$. As it is well known, each such g is a conformal automorphism of the Riemann sphere onto itself; in fact, it can be proved that these are the only ones.

Let **H** denote the upper half plane of $\mathbb{C}$, i.e.

$$\mathbf{H} = \{z \in \mathbb{C} \mid \text{Im } z > 0\}.$$

If $g \in \mathbf{SL}_2(\mathbb{R}) \subset \mathbf{SL}_2(\mathbb{C})$ then

$$\text{Im } g(z) = \frac{\text{Im } z}{|cz+d|^2}, \tag{1}$$

so these transformations fix **H**. Such a g has at least one, at most two fixed points; any such transformation is one of the following types:

- *Elliptic transformation* – this transformation has two fixed points ζ, $\bar{\zeta}$ with $\zeta \in \mathbf{H}$. After a change of variable sending $(\zeta, \bar{\zeta})$ to $(0, \infty)$, we get the following normal form:

  $$g(z) = Kz, \quad K = e^i\theta,$$

  i.e. it is a rotation about 0 through an angle θ.

- *Hyperbolic transformation* – this transformation has two fixed points on the real axis. By a suitable change of variable we can send them to $(0, \infty)$ and we get the normal form

  $$g(z) = Kz, \quad K > 0,$$

  i.e. it is a dilatation of magnitude K centered at the origin.

- *Parabolic transformation* – this has one fixed point on the real line, which can be sent to ∞ and we get

  $$g(z) = z + c$$

  which is a translation.

For an arbitrary matrix $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}_2(\mathbb{R})$, c=0 implies that g is parabolic, otherwise
$|a + d| > 2 \Longleftrightarrow g$ hyperbolic,
$|a + d| = 2 \Longleftrightarrow g$ parabolic,
$|a + d| < 2 \Longleftrightarrow g$ elliptic.
(The assertions in the last few paragraphs can be checked easily by direct calculation.)

Now consider the subgroup $\Gamma' = \mathbf{SL}_2(\mathbb{Z}) \subset \mathbf{SL}_2(\mathbb{R})$. This group acts discontinuously on $\mathbf{H}$, it is called the *homogeneous modular group*. The element $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \in \Gamma'$ acts trivially on $\mathbf{H}$ so we consider $\Gamma = \Gamma'/\pm\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ the *inhomogeneous modular group* or simply modular group.

For the remainder of this essay we shall denote

$$\mathbf{S} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{T} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

as elements of Γ, and also the corresponding transformations of $\mathbf{H}$:

$$\mathbf{S}(z) = -\frac{1}{z}, \quad \mathbf{T}(z) = z + 1.$$

If g is an elliptic transformation of Γ then by the above $a+b = 0$ or $a+b = \pm 1$. A simple calculation shows that the first implies $g^2 = \mathrm{I}$, whereas the second implies $g^3 = \mathrm{I}$, where I is the identity of Γ. Points fixed by such transformations will be called *elliptic fixed points*, they have periods 2 or 3. E.g. $i = \sqrt{-1}$ is fixed by $\mathbf{S}$ of order 2, $\rho = \frac{-1+\sqrt{3}i}{2}$ is fixed by $\mathbf{ST}$ of order 3. Similarly, points fixed by parabolic transformations will be called *parabolic fixed points*, they always belong to $\mathbb{Q} \cup \{\infty\}$.

Two points $z_1, z_2 \in \mathbf{H}$ are called *equivalent* under a group G of transformations of $\mathbf{H}$, written $z_1 \sim_G z_2$, if there is a transformation $g \in G$ with $g(z_1) = z_2$. This is clearly an equivalence relation, orders the points of $\mathbf{H}$ into equivalence classes. A *fundamental domain* for the action of G on $\mathbf{H}$ is an open, connected subset $D \subset H$ with the following properties:

- D does not contain any pair of distinct equivalent points,
- the point set closure of D contains at least one point from each equivalence class.

Now we claim, that the familiar region $\mathbf{D} = \{z \in \mathbf{H} : |z| > 1, |\operatorname{Re} z| < 1/2\}$ is a fundamental region for the action of Γ on $\mathbf{H}$. (See Figure 1.) This will be clear from the following

Theorem 1. (i) For every $z \in \mathbf{H}$ there exists $g \in \Gamma$ such that $g(z) \in \bar{\mathbf{D}}$ (the closure of $\mathbf{D}$).

(ii) If two distinct points $z_1, z_2 \in \bar{\mathbf{D}}$ are equivalent then $Re z_1 = \pm 1/2$ and $z_2 = \mathbf{T}^{\pm 1}(z_1)$, or $|z_1| = 1$ and $z_2 = \mathbf{S}(z_1)$.

(iii) Let $\Gamma_z = \{g : \ g \in \Gamma, g(z) = z\}$ be the stabilizer of z in Γ. Then $\Gamma_z = \{\mathrm{I}\}$ except for the following three cases:

$z \sim_\Gamma i$, in which case Γ_z is conjugate to $\langle \mathbf{S} \rangle$ of order 2,

$z \sim_\Gamma \rho$, in which case Γ_z is conjugate to $\langle \mathbf{ST} \rangle$ of order 3,

$z \sim_\Gamma -\bar{\rho}$, in which case Γ_z is conjugate to $\langle \mathbf{TS} \rangle$ of order 3.

The order of the stabilizer of a point P$\in \mathbf{H}$ will be denoted by e_P.

Proof: The proof of this theorem is not very complicated, but somewhat lengthy, so we do not repeat it here – it can be found in Serre [6] on page 78. □

As a by-product of the proof, we also get the following important fact:

Theorem 2. The group Γ is generated by $\mathbf{S}$ and $\mathbf{T}$.

□

2.2 Analytic structure and compactification

Consider $\mathbf{H}/\Gamma$, the set of equivalence classes of $\mathbf{H}$ – we want to put a topological and an analytic structure on it. We topologize $\mathbf{H}/\Gamma$ with the strongest topology

Figure 1: The fundamental domain of Γ

under which the natural map $\tau : \mathbf{H} \longrightarrow \mathbf{H}/\Gamma$ is continuous – $\mathbf{H}/\Gamma$ is simply $\bar{D}$ with proper identifications along the boundary as in Theorem 1.

As far as the analytic structure is concerned, around any point z in $\bar{\mathbf{D}}$ which is not a fixed point, we can take a small disc which is mapped conformally onto an open neighbourhood of $\tau z \in \mathbf{H}/\Gamma$ – this defines a parametric disc around τz. However, the elliptic fixed points must be treated separately. By Theorem 1, there are two equivalence classes of such points, i and ρ. Near i we can take $\tilde{z} = \left(\frac{z-i}{z+i}\right)^2$ as a local parameter (take a "half-disc" around i with the two radii identified by $\mathbf{S}$, transform $\mathbf{S}$ into normal form and send each element to its square to get a full disc as a parametric disc), whereas around ρ we can take $\tilde{z} = \left(\frac{z-\rho}{z+\bar{\rho}}\right)^3$ for similar reasons.

We can compactify $\mathbf{H}/\Gamma$ by adding the point ∞. To get an analytic structure on the compactification, we must find a parametric disk around ∞. But the set $\{z \in \mathbf{H} : \text{Im } z > 1\}$ is mapped by $t : z \longrightarrow e^{2\pi i z}$ onto the punctured disc $|t| < e^{-2\pi}$. For fixed x, as $y \to \infty$, $\arg(t(z))$ remains constant, while $|t(z)|$

approaches 0. Finally points identified by t differ by an integer m hence they are the same in $\mathbf{H}/\Gamma$. So we can add the point $t = 0$ in this parametric disc and use $q = e^{2\pi iz}$ as a local parameter around ∞. (This is sometimes called the *horocycle topology*.) It is also true, that this topological space is a Haussdorf space. This is not difficult, though nontrivial; the proof, together with the whole compactification procedure in detail, can be found in Shimura's book [7].

The resulting object is a compact Riemann surface which we shall denote by Π. There is a natural triangulation of $\mathbf{H}$ where the vertices are elliptic or parabolic fixed points, and the edges are images of the boundary of $\mathbf{D}$ under Γ (see Figure 1.) – this gives a natural triangulation for Π. From this triangulation one easily sees that the surface Π is a sphere.

Collecting our observations together, we get

Theorem 3. The space $\mathbf{H}/\Gamma$, compactified by adding the point ∞, can be given a natural analytic structure, under which it becomes a compact Riemann surface Π of genus 0.

□

2.3 Subgroups of Γ

Let $\mathrm{G} \subseteq \Gamma$ be a subgroup of the modular group of finite index μ. We shall find a fundamental domain for G, which can be compactified and made into a Riemann surface just as in 2.2.

Theorem 4. Let G be as above, and select coset representatives $\mathbf{T}_1, \ldots, \mathbf{T}_\mu$ so that $\Gamma = \mathrm{G}\mathbf{T}_1 \cup \mathrm{G}\mathbf{T}_2 \cup \ldots \cup \mathrm{G}\mathbf{T}_\mu$. If $\mathbf{D}$ is a fundamental domain for Γ then

$$\mathbf{D}_\mathrm{G} = \mathbf{T}_1\mathbf{D} \cup \mathbf{T}_2\mathbf{D} \cup \ldots \cup \mathbf{T}_\mu\mathbf{D}$$

is a fundamental domain for G.

Proof: The transforms of $\bar{\mathbf{D}}_\mathrm{G}$ by elements of G clearly cover $\mathbf{H}$, so the second property is satisfied. For the first property, if $g\mathbf{D}_\mathrm{G} \cap \mathbf{D}_\mathrm{G}$ would contain an open set for any $g \in \mathrm{G}$, that set would contain a transform of $\mathbf{D}$. But then $g\mathbf{T}_i\mathbf{D} = \mathbf{T}_j\mathbf{D}$ would imply $g\mathbf{T}_i = \mathbf{T}_j$, a contradiction since the $\mathbf{T}_i$ are representatives of distinct cosets. □

Now the quotient space $\mathbf{H}/\mathrm{G}$ can be given an analytic structure, just as $\mathbf{H}/\Gamma$ was in 2.2. As for the compactification, the local parameter at ∞ will be $q = e^{2\pi iz/m}$ where m is the smallest positive integer such that the transformation $z \to z + m$ is in G. There may also be real parabolic fixed points as well, but they can be treated in the same way – first sending them to ∞ by a suitable transformation and then using q as a local parameter there. The parabolic fixed points, finite or infinite, are also called *cusps* of the surface. As a result, we get a compact Riemann surface again, which we shall denote by Π_G.

The natural triangulation of $\mathbf{H}$ mentioned above induces a triangulation on Π_{G}, in which the fixed points are the vertices and every 1-simplex connects two fixed points. As it is well known, we can then compute the genus of Π_{G} by means of the Euler characteristic formula

$$\chi = 2 - 2p = \sigma_0 - \sigma_1 + \sigma_2,$$

where χ is the Euler characteristic, p is the genus and σ_k is the number of k-simpleces in the triangulation. The knowledge of the genus p is of special importance in applications of the Riemann-Roch Theorem at later stages of the theory, see Chapter 5.

This topic is discussed further in Gunning [1] in §4 with some specific examples. Here we only mention one class of subgroups, which are important in applications: the *congruence subgroups*. Γ'_n, the full homogeneous congruence subgroup of level n, where $n \geq 2$ is an integer, is defined by

$$\Gamma'_n = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma' : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \mathrm{I} \,(\bmod \mathrm{n}) \right\}.$$

Similarly the full inhomogeneous congruence subgroup of level n is defined by

$$\Gamma_n = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma' : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \pm\mathrm{I} \,(\bmod \mathrm{n}) \right\} / \{\pm \mathrm{I}\}.$$

More generally, any group $\Gamma_n \subseteq \mathrm{G} \subseteq \Gamma$ is called a congruence subgroup of level n.

The full congruence subgroups are studied in detail in Gunning [1] and Husemöller [2], we just cite the most important results. Denote

$$\nu'(n) = [\Gamma' : \Gamma'_n], \quad \nu(n) = [\Gamma : \Gamma_n].$$

Then for $n = 2$, $\nu'(2) = \nu(2)$ since $\mathrm{I} \equiv -\mathrm{I} \,(\bmod 2)$, whereas for $n > 2$, we have $\nu(n) = \frac{1}{2}\nu'(n)$.

A pair of integers (c, d) is called *primitive* mod n, if $(c, d, n) = 1$ using the usual notation for the greatest common divisor; let the number of incongruent primitive pairs mod n be $\lambda(n)$. An easy argument shows that $\Gamma'/\Gamma'_n \cong \mathbf{SL}_2(\mathbb{Z}_{\mathrm{n}})$ (the homomorphism $\tilde{\phi} : \mathbf{SL}_2(\mathbb{Z}) \to \mathbf{SL}_2(\mathbb{Z}_{\mathrm{n}})$, with kernel Γ'_n induced by the natural homomorphism $\phi : \mathbf{Z} \to \mathbf{Z}_n$ is onto) so

$$\nu'(n) = \mid \mathbf{SL}_2(\mathbb{Z}_{\mathrm{n}}) \mid = n\lambda(n),$$

where the last equality also follows from a simple number-theoretic argument. The Chinese Remainder Theorem implies that λ is multiplicative, i.e. if n_1 and n_2 are relatively prime, then $\lambda(n_1 n_2) = \lambda(n_1)\lambda(n_2)$. Finally, for a prime q

$$\lambda(q^k) = q^{2k}\left(1 - \frac{1}{q^2}\right).$$

Collecting all these observations together, we get

$$\nu(2) = 6, \quad \nu(n) = \frac{1}{2} n^3 \prod_{q|n} \left(1 - \frac{1}{q^2}\right) \quad \text{if } n > 2,$$

which can be used to compute the genus p of the corresponding surfaces Π_G as in the references above.

We note one more interesting fact. It can be checked that for $G = \Gamma_2$, Γ_3, Γ_4, Γ_5 the surface Π_G has genus 0, so it is a sphere. The groups Γ/Γ_n act on the sphere as groups of conformal automorphisms – they act as the dihedral group of order 6, the tetrahedral group, the octahedral group and the icosahedral group respectively.

3 Modular forms – the elementary approach

3.1 Modular functions and modular forms

Let k be an integer. A function f is *weakly modular* of weight 2k, if f is meromorphic on $\mathbf{H}$ and satisfies

$$f(z) = (cz + d)^{-2k} f\left(\frac{az + b}{cz + d}\right) \text{ for all } z \to \frac{az + b}{cz + d} \text{ belonging to } \Gamma. \qquad (2)$$

By Theorem 2, f is weakly modular of weight $2k$ if and only if f satisfies the two relations

$$\begin{array}{lllllll} f(\mathbf{T}z) & = & f(z) & \text{i.e.} & f(z+1) & = & f(z), \\ f(\mathbf{S}z) & = & z^{2k} f(z) & \text{i.e.} & f(-1/z) & = & z^{2k} f(z). \end{array} \qquad (3)$$

However, these conditions are not sufficent for our purposes. Informally, we want our functions f to be meromorphic also "at the cusp ∞". The local parameter there is $q = e^{2\pi i z}$; let

$$\tilde{f}(q) = f(z)$$

i.e. we express f in terms of the variable q. Then $\tilde{f}$ will be meromorphic in the disk $| q | < 1$ with the origin removed. If $\tilde{f}$ extends to a meromorphic (resp. holomorphic) function at the origin, we say f is meromorphic (resp. holomorphic) at infinity. If f is holomorphic at infinity, we set $f(\infty) = \tilde{f}(0)$.

Now a weakly modular function f of weight $2k$, which is meromorphic also at the cusp ∞, is called a *modular function* of the same weight. A modular function which is holomorphic everywhere (including infinity) is called a *modular form* of weight $2k$. Finally, if $f(\infty) = 0$ then we call f a *cusp form* of weight $2k$.

3.2 Connection with lattice functions

Suppose $\mathbf{V}$ is a finite dimensional real vector space. Recall that a lattice $\mathcal{L}$ is a subgroup of $\mathbf{V}$ satisfying one of the following equivalent conditions:

- $\mathcal{L}$ is discrete and $\mathbf{V}/\mathcal{L}$ is compact;
- $\mathcal{L}$ is discrete and generates the $\mathbb{R}$-vector space $\mathbf{V}$;
- there exists an $\mathbb{R}$-basis $(e_1, e_2, \ldots, e_n)$ of $\mathbf{V}$ which is a $\mathbb{Z}$-basis of $\mathcal{L}$.

Let $\mathcal{R}$ be the set of lattices of $\mathbb{C}$ (considered as a 2-dimensional $\mathbb{R}$-space), and let M be the set of pairs (ω_1, ω_2) of elements of $\mathbb{C}^*$ with Im $(\omega_1/\omega_2) > 0$. To each such pair we associate a lattice

$$\mathcal{L}(\omega_1, \omega_2) = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$$

with basis $\{\omega_1, \omega_2\}$. If $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}_2(\mathbb{Z})$, and

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

then $\{\omega_1', \omega_2'\}$ is also a basis for $\mathcal{L}(\omega_1, \omega_2)$ with $\frac{\omega_1'}{\omega_2'} = g\left(\frac{\omega_1}{\omega_2}\right)$, so Im $(\omega_1'/\omega_2') > 0$, i.e. $(\omega_1', \omega_2') \in M$.

Let now F be a complex-valued function on $\mathcal{R}$, and let $k \in \mathbb{Z}$. We say F is of order $2k$ if

$$F(\lambda\mathcal{L}) = \lambda^{-2k} F(\mathcal{L}) \tag{4}$$

for all lattices $\mathcal{L}$ and all $\lambda \in \mathbb{C}^*$.

If F is such a function and $(\omega_1, \omega_2) \in M$, we denote by $F(\omega_1, \omega_2)$ the value of F on the lattice $\mathcal{L}(\omega_1, \omega_2)$. Formula (4) translates to

$$F(\lambda\omega_1, \lambda\omega_2) = \lambda^{-2k} F(\omega_1, \omega_2), \tag{5}$$

and also $F(\omega_1, \omega_2)$ is invariant under $\mathbf{SL}_2(\mathbb{Z})$.

Formula (5) shows that the product $\omega_2^{2k} F(\omega_1, \omega_2)$ depends only on $z = \frac{\omega_1}{\omega_2}$. So there exists a function f on $\mathbf{H}$ (remember $(\omega_1, \omega_2) \in M$), such that

$$F(\omega_1, \omega_2) = \omega_2^{-2k} f(\omega_1/\omega_2). \tag{6}$$

But F is invariant under $\mathbf{SL}_2(\mathbb{Z})$, so f satisfies the identity

$$f(z) = (cz + d)^{-2k} f\left(\frac{az + b}{cz + d}\right) \text{ for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}_2(\mathbb{Z}) \tag{7}$$

i.e. f is a modular function. Conversely, if f is such a function, then formula (6) associates to it a lattice function F. Thus we can identify modular functions of weight $2k$ with lattice functions of weight $2k$.

3.3 Eisenstein series

We begin with a quick
Lemma Let $\mathcal{L}$ be a lattice in $\mathbb{C}$. The series

$$\sum_{\gamma\in\mathcal{L}}{}' \frac{1}{|\gamma|^{\sigma}}$$

is convergent for $\sigma > 2$. (Here and in what follows $\sum'$ indicates that the summation is over all nonzero elements.)

Proof: The number of elements of $\mathcal{L}$ such that $|\gamma|$ is between two consecutive integers n and $n+1$ is $O(n)$. So the series can be estimated from above by a multiple of $\sum 1/n^{\sigma-1}$ which converges for $\sigma > 2$.

□

Let now $k > 1$ be an integer. Put

$$G_k(\mathcal{L}) = \sum_{\gamma\in\mathcal{L}}{}' \frac{1}{\gamma^{2k}}.$$

The series converges absolutely by the above Lemma. Clearly G_k is of weight $2k$, it is called the *Eisenstein series* of weight $2k$. We can view G_k as a function on M, given by

$$G_k(\omega_1,\omega_2) = \sum_{m,n}{}' \frac{1}{(m\omega_1+n\omega_2)^{2k}}.$$

Here again, $\sum'$ means that summation is over all pairs distinct from $(0,0)$. Finally, this gives a function on $\mathbf{H}$, also denoted G_k:

$$G_k(z) = \sum_{m,n}{}' \frac{1}{(mz+n)^{2k}}. \tag{8}$$

Theorem 5. The Eisenstein series $G_k(z)$, for $k > 1$ an integer, is a modular form of weight $2k$. We have $G_k(\infty) = 2\zeta(2k)$, where ζ denotes the Riemann zeta function.

Proof: The above argument shows that $G_k(z)$ is weakly modular of weight $2k$. We have to show that it is holomorphic everywhere including infinity. First suppose that z is in the fundamental domain $\mathbf{D}$. Then

$$\begin{aligned} |mz+n|^2 &= m^2 z\bar{z} + 2mnRe(z) + n^2 \\ &\geq m^2 - mn + n^2 \\ &= |m\rho+n|^2 \end{aligned}$$

and by our Lemma the series $\sum' 1/\,|\,m\rho+n\,|\,^{2k}$ is convergent. This shows that the series $G_k(z)$ converges uniformly on $\mathbf{D}$, thus also in the transforms of $\mathbf{D}$; but these transforms cover $\mathbf{H}$, so $G_k(z)$ is holomorphic on $\mathbf{H}$.

Finally, we have to prove that G_k has a limit as Im $z \to \infty$. One may suppose that z remains in $\mathbf{D}$, and by uniform convergence there, one can take the limit termwise. The terms $(mz+n)^{-2k}$ with $m \neq 0$ give 0, the others give n^{-2k} thus indeed

$$G_k(\infty) = \sum{}' \frac{1}{n^{2k}} = 2\zeta(2k).$$

□

Because of the theory of elliptic curves, one usually replaces G_2 and G_3 by multiples $g_2 = 60G_2$, $g_3 = 140G_3$. Using the known values of $\zeta(4)$ and $\zeta(6)$, one gets

$$g_2(\infty) = \frac{4}{3}\,\pi^4, \quad g_3(\infty) = \frac{8}{27}\,\pi^6.$$

If we put

$$\Delta = g_2^3 - 27g_3^2, \tag{9}$$

then Δ is a modular form of weight 12 with $\Delta(\infty) = 0$, i.e. Δ is a cusp form.

3.4 The space of modular forms

For k an integer, denote by M_k (resp. M_k^0) the $\mathbb{C}$-vector space of modular forms of weight $2k$ (resp. cusp forms of weight $2k$). M_k^0 is the kernel of the linear form $f \to f(\infty)$ on M_k, thus $\dim M_k/M_k^0 \leq 1$. But for $k \geq 2$, the Eisenstein series G_k is an element of M_k with $G_k(\infty) \neq 0$, hence

$$M_k = M_k^0 \oplus \mathbb{C}.G_k \text{ for } k \geq 2.$$

In studying the space of modular forms, the following Theorem will be our basic tool:

If f is a meromorphic function on $\mathbf{H}$ which is not identically 0 and P$\in \mathbf{H}$, one can find an integer n such that $f/(z-P)^n$ is holomorphic and nonzero near P. n is called the order of f at P, denoted $v_P(f)$. When f is a modular function, our basic identity (2) shows that $v_P(f) = v_{g(P)}(f)$ if $g \in \Gamma$, i.e. $v_P(f)$ depends only on the image of P in $\mathbf{H}/\Gamma$. One can also define $v_\infty(f)$ as the order of $\tilde{f}(q)$ at $q = 0$.

Theorem 6. If f is a modular function of weight $2k$, not identically 0, then

$$v_\infty(f) + \sum_{P \in H/\Gamma} \frac{1}{e_P} v_P(f) = \frac{k}{6}. \tag{10}$$

(Recall e_P is the order of the stabilizer of P.)

Proof: Observe first that f has only a finite number of zeros and poles modulo Γ. Indeed, since $\tilde{f}$ is meromorphic, there exists an $r > 0$ such that $\tilde{f}$ has no zero or pole for $0 < |\, q \,| < r$; this means that f has no zero or pole for $Im(z) > e^{2\pi r}$. On the other hand, $\mathbf{D} \cap \{z : \text{Im } (z) \leq e^{2\pi r}\}$ is compact, so f has only a finite number of poles and zeros there.

Suppose first that no zero or pole falls on the boundary of $\mathbf{D}$ except possibly at $i, \rho, -\bar{\rho}$. We will integrate $\frac{1}{2\pi i}\frac{df}{f}$ around the boundary of $\mathbf{D}$ modified so as to exclude the possible poles and zeros at $i, \rho, -\bar{\rho}$ but to include all other poles and zeros in $\mathbf{D}$. (See Figure 2.) We have by the residue theorem

$$\frac{1}{2\pi i}\int_{\mathcal{C}} \frac{df}{f} = \sum_{\substack{P \in H/\Gamma \\ P \neq i, \rho, -\bar{\rho}}} v_P(f).$$

Figure 2: The contour $\mathcal{C}$ of integration

On the other hand,
(a) the change of variables $q = e^{2\pi i z}$ transforms the arc EA into a circle C centered at $q = 0$ with negative orientation, not enclosing any zeros or poles except possibly at 0. Hence

$$\frac{1}{2\pi i}\int_A^E \frac{df}{f} = \frac{1}{2\pi i}\int_C \frac{df}{f} = -v_\infty(f).$$

(b) The integral of $\frac{1}{2\pi i}\frac{df}{f}$ on the circle containing BB', oriented negatively, has value $-v_\rho(f)$. When the radius of the circle tends to 0, the angle $B\rho B'$ tends to $\frac{2\pi}{6}$. Hence

$$\frac{1}{2\pi i}\int_B^{B'}\frac{df}{f} \to -\frac{1}{6}v_\rho(f).$$

Similarly

$$\frac{1}{2\pi i}\int_C^{C'}\frac{df}{f} \to -\frac{1}{2}v_i(f),$$

$$\frac{1}{2\pi i}\int_D^{D'}\frac{df}{f} \to -\frac{1}{6}v_{-\bar\rho}(f).$$

(c) The arc AB is transformed by $\mathbf{T}$ onto the arc ED'; since $f(\mathbf{T}z) = f(z)$, we get cancellation along these arcs.
(d) S transforms the arc B'C onto the arc DC'; since $f(\mathbf{S}z) = z^{2k}f(z)$, we get

$$\frac{df(\mathbf{S}z)}{f(\mathbf{S}z)} = 2k\frac{dz}{z} + \frac{df(z)}{f(z)},$$

hence

$$\begin{aligned}\frac{1}{2\pi i}\int_{B'}^{C}\frac{df}{f} + \frac{1}{2\pi i}\int_{C'}^{D}\frac{df}{f} &= \frac{1}{2\pi i}\int_{B'}^{C}\left(\frac{df(z)}{f(z)} - \frac{df(\mathbf{S}z)}{f(\mathbf{S}z)}\right)\\ &= \frac{1}{2\pi i}\int_{B'}^{C}\left(-2k\frac{dz}{z}\right)\\ &\to -2k\left(-\frac{1}{12}\right)\\ &= \frac{k}{6},\end{aligned}$$

when the radii of the small circles tend to 0.

Writing now that the two expressions for $\frac{1}{2\pi i}\int_{\mathcal{C}}\frac{df}{f}$ are equal and passing to the limit, we get formula (10).

If f has poles or zeros on the contour, we modify it slightly to remove them from $\mathcal{C}$ (e.g. if f has a pole at z_0 on the arc AB, then take a small half-circle around z_0 and $\mathbf{T}(z_0)$ so that one of them will be inside, the other outside $\mathcal{C}$) and repeat the above proof.

□

Theorem 7. (i) We have $M_k = 0$ for $k < 0$ and $k = 1$.

(ii) For $k = 0, 2, 3, 4, 5$, M_k is a vector space of dimension 1 with basis 1, G_2, G_3, G_4, G_5; for these k-s, we have $M_k^0 = 0$.

(iii) Δ is nonzero on $\mathbf{H}$ and has a simple zero at infinity.

(iv) Multiplication by Δ defines an isomorphism of M_{k-6} onto M_k^0.

(v) We have for $k \geq 0$

$$\dim M_k = \begin{cases} [k/6] & \text{if } k \equiv 1 \pmod 6, \\ [k/6] + 1 & \text{if } k \not\equiv 1 \pmod 6. \end{cases} \tag{11}$$

(vi) The space M_k has for basis the monomials $G_2^a G_3^b$, with a, b nonnegative integers satisfying $2a + 3b = k$.

Proof: We use our basic formula (10). All terms on the left hand side are nonnegative (f cannot have any poles), thus we have $k \geq 0$; also $k \neq 1$ since $\frac{1}{6}$ cannot be written in the form $n_1 + \frac{n_2}{2} + \frac{n_3}{3}$ with n_1, n_2, n_3 nonnegative integers.

Now apply (10) to $f = G_2$. $\frac{2}{6} = n_1 + \frac{n_2}{2} + \frac{n_3}{3}$ implies that $n_1 = 0$, $n_2 = 0$, $n_3 = 1$. Hence $v_\rho(G_2) = 1$, and G_2 is nonzero at other points. A similar argument shows that $v_i(G_3) = 1$, and G_3 is nonzero at other points. This already gives that Δ does not vanish identically, since it does not vanish at i.

The weight of Δ is 12, and $v_\infty(\Delta) \geq 1$, so (10) implies that $v_P(\Delta) = 0$ for $P \neq \infty$ and $v_\infty(\Delta) = 1$. This gives (iii).

If $f \in M_k^0$ and we set $g = f/\Delta$ then g will be a modular function of weight $2k-12$, but g will also be holomorphic on $\mathbf{H} \cup \{\infty\}$ (f is zero at ∞) so $g \in M_{k-6}$, and this is clearly a bijection between the two spaces, hence (iv).

For $k \leq 5$ we have $k - 6 < 0$, so by (i) and (iv) $M_k^0 = 0$; this shows that $\dim M_k \leq 1$. But 1, G_2, G_3, G_4, G_5 are nonzero elements of the corresponding M_k, hence (ii).

(i) and (ii) show that formula (11) is true if $0 \leq k < 6$. By (iv) both sides increase by 1, when k is replaced by $k + 6$. Hence by induction, (v) holds.

As far as (vi) is concerned, first we have to show that the monomials generate M_k. By (i) and (ii) this is clear if $k \leq 3$. For $k \geq 4$, one uses induction. If $f \in M_k$, choose a nonnegative integer pair (a, b) with $2a + 3b = k$ – this is always possible – and form $g = G_2^a G_3^b$ which is nonzero at ∞. For a suitable λ $f - \lambda g$ will be a cusp form, equal to $h\Delta$ for some $h \in M_{k-6}$ by (iv). Finally, one applies the inductive hypothesis to h.

To end the proof, we must show that the monomials are linearly independent. But if they were not, G_2^3/G_3^2 would be a root of a nonzero polynomial over $\mathbb{C}$, i.e. it would be constant by continuity. But this is clearly impossible (look at the values near i and ρ).

□

3.5 The modular invariant

Put

$$j = 1728\frac{g_2^3}{\Delta}. \tag{12}$$

Theorem 8. (i) The function j is a modular function of weight 0.
(ii) It is holomorphic in $\mathbf{H}$ and has a simple pole at infinity.
(iii) It defines a bijection of $\mathbf{H}/\Gamma$ onto $\mathbb{C}$.

Proof: The first statement is obvious from the definition. The second statement is a simple application of the properties of Δ given in Theorem 7.(iii). (g_2 is nonzero at infinity). Actually, the coefficient 1728 ensures that j has residue 1 at ∞, which is convenient in some calculations.

Now consider $f_\lambda = 1728g_2^3 - \lambda\Delta$. This is a modular form of weight $2k = 12$, so applying formula (10) we get $1 = n_1 + \frac{n_2}{2} + \frac{n_3}{3}$. But the only decompositions are $(1,0,0)$, $(0,2,0)$, $(0,0,3)$, i.e. f_λ has one zero on $\mathbf{H}/\Gamma$, which establishes (iii).

□

Theorem 9. Let f be a meromorphic function on $\mathbf{H}$. The following are equivalent:
(i) f is a modular function of weight 0,
(ii) f is a quotient of two modular forms of the same weight,
(iii) f is a rational function of j.

Proof: Clearly (iii)⇒(ii)⇒(i), so we must show (i)⇒(iii). Suppose f is a function satisfying (i). We can multiply f by a suitable power of j to ensure, that f is holomorphic on $\mathbf{H}$. Since $\Delta(\infty) = 0$, there exists an n such that $g = \Delta^n f$ is holomorphic also at infinity. g is then a modular form of weight $12n$, so by Theorem 7.(vi) we can write it as a linear combination of the $G_2^a G_3^b$ with $3a + 2b = 6n$. By linearity, it suffices to consider one such term, i.e. $f = G_2^a G_3^b/\Delta^n$. But from $3a+2b = 6n$ we see that $c = a/2$, $d = b/3$ are integers and this gives $f = \left(G_2^3/\Delta\right)^c \left(G_3^2/\Delta\right)^d$. Finally, G_2^3/Δ and G_3^2/Δ are clearly rational functions of j.

□

We remark here, that Theorem 8.(iii) establishes an explicit isomorphism between the Riemann sphere $\mathbb{C} \cup \{\infty\}$ and the surface Π which is also a sphere (see Chapter 2.2). Theorem 9.(iii) is equivalent to the well known fact that the only meromorphic functions on the Riemann sphere are the rational functions; j was shown to be a generator of that function field.

3.6 Expansions at infinity

Suppose f is a modular form of some weight $2k$. Then the corresponding function $\tilde{f}$ is holomorphic at infinity, it has a power series there:

$$\tilde{f}(q) = \sum_{n=0}^{\infty} a_n q^n.$$

In terms of the variable z this yields

$$f(z) = \sum_{n=0}^{\infty} a_n e^{2\pi i n z},$$

i.e. f has a Fourier expansion; the a_n are called the Fourier coefficients of f. In this section we want to give the Fourier expansion of some of the forms considered sofar, and give estimates in the general case. With some abuse of notation, we will change between the variables z and q freely even in one formula, but this will make no confusion. Denote

$$\sigma_k(n) = \sum_{d|n} d^k$$

the sum of kth powers of positive divisors of n. We also denote by B_k the kth Bernoulli number; they are defined by

$$z \cot z = 1 - \sum_{k=1}^{\infty} B_k \frac{2^{2k}}{(2k)!} z^{2k},$$

and satisfy the indentity

$$\zeta(2k) = \frac{2^{2k-1}}{(2k)!} \pi^{2k} B_k \tag{13}$$

(see Serre [6]). The first few values are given by

$$B_1 = \frac{1}{6}, \quad B_2 = \frac{1}{30}, \quad B_3 = \frac{1}{42}, \quad B_4 = \frac{1}{30},$$

$$B_5 = \frac{5}{66}, \quad B_6 = \frac{691}{2730}, \quad B_7 = \frac{7}{6}.$$

Theorem 10. For $k \geq 2$ we have

(i) $G_k(z) = 2\zeta(2k) + \dfrac{2}{(2k-1)!} (2i\pi)^{2k} \displaystyle\sum_{n=1}^{\infty} \sigma_{2k-1}(n) q^n.$

(ii) $G_k(z) = 2\zeta(2k) E_k(z)$ with

$$E_k(z) = 1 + \gamma_k \sum_{n=1}^{\infty} \sigma_{2k-1}(n) q^n \tag{14}$$

and

$$\gamma_k = (-1)^k \frac{4k}{B_k}.$$

(iii) The Fourier coefficients a_n of G_k satisfy

$$An^{2k-1} \leq |\, a_n \,| \leq Bn^{2k-1} \tag{15}$$

with positive constants A, B.

Proof: (sketched)

We start with the known formula

$$\pi \cot \pi z = \frac{1}{z} + \sum_{m=1}^{\infty} \left(\frac{1}{z+m} + \frac{1}{z-m} \right).$$

On the other hand

$$\pi \cot \pi z = \pi \frac{\cos \pi z}{\sin \pi z} = i\pi \frac{q+1}{q-1} = i\pi - \frac{2i\pi}{1-q} = i\pi - 2i\pi \sum_{n=0}^{\infty} q^n.$$

Comparing, we get

$$\frac{1}{z} + \sum_{m=1}^{\infty} \left(\frac{1}{z+m} + \frac{1}{z-m} \right) = i\pi - 2i\pi \sum_{n=0}^{\infty} q^n.$$

By successive differentiations with respect to z, we have for $k \geq 2$

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

Using this expression, after some calculations

$$\begin{aligned} G_k(z) &= {\sum_{(m,n)}}' \frac{1}{(nz+m)^{2k}} = 2\zeta(2k) + 2 \sum_{n=1}^{\infty} \sum_{m \in Z} \frac{1}{(nz+m)^{2k}} \\ &= 2\zeta(2k) + \frac{2}{(2k-1)!} (2i\pi)^{2k} \sum_{n=1}^{\infty} \sigma_{2k-1}(n) q^n, \end{aligned}$$

hence (i). (ii) is also clear, we only need to calculate the coefficient γ_k. Using (13)

$$\gamma_k = \frac{(2\pi i)^{2k}}{(2k-1)!} \frac{(2k)!}{2^{2k-1} \pi^{2k} B_k} = (-1)^k \frac{4k}{B_k}.$$

Now (ii) shows that there exists a constant A such that

$$a_n = (-1)^k A \sigma_{2k-1}(n),$$

hence

$$|\, a_n \,| = A\sigma_{2k-1}(n) \geq An^{2k-1}.$$

On the other hand,

$$\frac{|\, a_n \,|}{n^{2k-1}} = A \sum_{d|n} \frac{1}{d^{2k-1}} \leq A \sum_{d=1}^{\infty} \frac{1}{d^{2k-1}} < \infty$$

since $k \geq 2$, hence (iii).

□

For further reference, we spell out some of the expansions given above:

$$\begin{aligned} E_2(q) &= 1 + 240 \textstyle\sum_{n=1}^{\infty} \sigma_3(n) q^n, \\ E_3(q) &= 1 - 504 \textstyle\sum_{n=1}^{\infty} \sigma_5(n) q^n, \\ E_4(q) &= 1 + 480 \textstyle\sum_{n=1}^{\infty} \sigma_7(n) q^n, \\ E_5(q) &= 1 - 264 \textstyle\sum_{n=1}^{\infty} \sigma_9(n) q^n, \\ E_7(q) &= 1 - 24 \textstyle\sum_{n=1}^{\infty} \sigma_{13}(n) q^n. \end{aligned} \tag{16}$$

Theorem 11. (i) If $f(z) = \sum_{n=1}^{\infty} a_n q^n$ is a cusp form of weight $2k$, then

$$a_n = O\,(n^k). \tag{17}$$

(ii) If $g(z) = \sum_{n=0}^{\infty} b_n q^n$ is not a cusp form, then

$$b_n = O\,(n^{2k-1}). \tag{18}$$

Proof: Put $y = Im(z)$ and $\phi(z) = |\, f(z) \,|\, y^k$. Then formulas (1) and (2) show that ϕ is invariant under Γ. Also ϕ is continuous on $\mathbf{D}$, and

$$|\, f(z) \,| = O\,(q) = O\,(e^{-2\pi y}),$$

hence $\phi \to 0$ as $y \to \infty$. This shows that ϕ is bounded, i.e. there is a positive M with

$$|\, f(z) \,| \leq My^{-k} \quad \text{for } z \in \mathbf{H}. \tag{19}$$

If we now fix y and vary $x = \text{Re } z$ between 0 and 1, then $q = e^{2\pi i(x+iy)}$ runs along a circle $\mathcal{C}$ centered at 0, hence by the residue formula

$$a_n = \frac{1}{2\pi i} \int_{\mathcal{C}} f(z) q^{-n-1} dq = e^{2\pi ny} \int_0^1 f(x+iy) e^{-2\pi inx}\, dx.$$

Using (19), we get

$$|a_n| \leq My^{-k}e^{2\pi ny}$$

valid for $y > 0$. Putting $y = \frac{1}{n}$ yields (i).

If now g is not a cusp form, we have $g = \lambda G_k + f$ with f a cusp form and $\lambda \neq 0$. (ii) now follows from (i) and the previous Theorem, since n^k is negligible compared to n^{2k-1}.

□

We note here, that the best possible bound is given by the

Ramanujan Conjectures 1: If $f(z) = \sum_{n=1}^{\infty} a_n q^n$ is a cusp form of weight $2k$, then

$$a_n = O(n^{k-1/2+\varepsilon}) \quad \text{for all } \varepsilon > 0.$$

Ramanujan in his famous paper [4] set a special case of this conjecture first, together with the other conjecture below. This one was only solved in 1973 by Deligne.

Finally we give one more expansion, which and the related questions had a strong influence on the development of the theory of modular forms.

Theorem 12. $\Delta = (2\pi)^{12} q \prod_{n=1}^{\infty} (1-q^n)^{24}$.

We do not prove this here - an elementary proof can be found in Serre [6], another approach is presented in Sarnak [5] in Appendix 1.1.

If now one denotes

$$q \prod_{n=1}^{\infty} (1-q^n)^{24} = \sum_{n=1}^{\infty} \tau(n) q^n,$$

then we can state

Ramanujan Conjectures 2: $\tau(n)$ is a multiplicative function, i.e.

$$\tau(mn) = \tau(m)\tau(n), \quad \text{if } (m,n) = 1.$$

This was first solved by Mordell; it can be proved using the fact, that Δ turns out to be an eigenfunction of a certain class of operators on M_6^0 called the Hecke operators. We do not have time to introduce them here, but the details can again be found in Serre [6].

4 Connection with elliptic curves

In this chapter we want to establish shortly the connection between elliptic curves and modular forms. The reference to this part of the essay is Husemöller [2]; the omitted proofs can be found there.

4.1 Elliptic Curves

An *elliptic curve* E is a projective nonsingular cubic curve over a field k, together with a commutative group law defined on the points of E. By projective changes of variables, the equation of E can always be brought in the form

$$Y^2Z = 4X^3 - g_2XZ^2 - g_3Z^3, \tag{20}$$

or in affine coordinates

$$y^2 = 4x^3 - g_2x - g_3.$$

To E we associate two quantities defined by

$$\Delta(E) = g_2^3 - 27g_3^2, \qquad j(E) = \frac{g_2^3}{\Delta(E)}.$$

Theorem 13. The equation (20) defines an elliptic curve over $\mathbb{C}$ if and only if $\Delta(E) = 0$. Two elliptic curves E and E' are isomorphic if and only if $j(E) = j(E')$. Moreover, if $j \in \mathbb{C}$ then $j = j(E)$ for some elliptic curve E.

□

4.2 The Weierstrass equation

Let $\mathcal{L} = \mathcal{L}(\omega_1, \omega_2)$ be a lattice in $\mathbb{C}$ as in 3.2. We have a quotient space $T = \mathbb{C}/\mathcal{L}$, which is topologically a torus, with a commutative group law inherited from the usual addition on $\mathbb{C}$. We associate to this lattice the Weierstrass function

$$\wp(u;\mathcal{L}) = \frac{1}{u^2} + \sum_{\gamma\in\mathcal{L}}{}' \left[\frac{1}{(u-\gamma)^2} - \frac{1}{\gamma^2}\right].$$

(Remember $\sum'$ denotes that we omit 0 from the summation.) The sum behaves like $1/\gamma^3$ for $|\gamma|$ large, so by our Lemma in 3.2, it converges locally uniformly on $\mathbb{C} - \mathcal{L}$, and defines a meromorphic function on $\mathbb{C}$ with a double pole at each $\gamma \in \mathcal{L}$. This function is an example of a class of functions called *elliptic functions*, i.e. the class of meromorphic functions on $\mathbb{C}$, which are invariant under translations by elements of the lattice $\mathcal{L}$.

After a straightforward computation, we get the expansion

$$\wp(u;\mathcal{L}) = \frac{1}{u^2} + \sum_{k=2}^{\infty}(2k-1)G_k(\mathcal{L})u^{2k-2},$$

where $G_k(\mathcal{L})$ denotes the Eisenstein series introduced in 3.3.

Using this expansion, taking the derivative of $\wp(u;\mathcal{L})$ with respect to u and comparing coefficients, we get the famous Weierstrass differential equation for $\wp$:

$$\wp'(u;\mathcal{L})^2 = 4\wp(u;\mathcal{L})^3 - 60G_2(\mathcal{L})\wp(u;\mathcal{L}) - 140G_3(\mathcal{L}).$$

It is now an easy thing to prove, that the map $h : \mathbb{C}/\mathcal{L} \to E(\mathcal{L})$, where $E(\mathcal{L})$ is the elliptic curve over $\mathbb{C}$ defined by

$$Y^2Z = 4X^3 - 60G_2(\mathcal{L})XZ^2 - 140G_3(\mathcal{L})Z^3,$$

and $h(u \bmod \mathcal{L}) = (\wp(u;\mathcal{L}),\ \wp'(u,\mathcal{L}),\ 1)$; $h(0 \bmod \mathcal{L}) = (0,\ 1,\ 0)$ is an analytic group isomorphism. This shows the strong connection between elliptic curves and complex tori (and that the additive structure of the torus carries over to $E(\mathcal{L})$ to give the commutative structure there), but it is more important for us, that for a curve $E(\mathcal{L})$ the constants g_2, g_3 of equation (20) are just the modified Eisenstein functions in the variable $z = \omega_1/\omega_2$ introduced in 3.3. Then the Δ and j functions are also the same, and in view of Theorem 13, we get an independent proof of the fact that Δ is nonzero on $\mathbf{H}$, whereas j takes all complex values on $\mathbf{H}$.

5 Modular forms on the surface Π_G

5.1 Functions and differentials on a compact Riemann surface

In this section we give a brief account of the most important facts concerning functions and differentials on an arbitrary compact Riemann surface Π_0. The proofs not given here can be found in Springer [8].

$f(z)$ is a meromorphic function on Π_0, if f has a Laurent expansion in the local coordinate z around any point P of Π_0:

$$f(z) = z^n \sum_{m=0}^{\infty} a_m z^m$$

with $a_0 \neq 0$, n an integer. We define n to be the order of f at P, written $v_P(f) = n$; the $\mathbb{C}$-vector space of meromorphic functions on Π_0 will be denoted by $\mathcal{M}$.

A (meromorphic) *differential form* ω on the surface is a correspondence, which associates to each point P of the surface, and each local parameter z at P a meromorphic function $g(z)$ with $\omega = g(z)dz$, such that if t is another local parameter at P with $\omega = h(t)dt$ then $g(z) = \frac{dt}{dz}h(t)$. If $\omega = g(z)dz$ is a differential form, then we denote $v_P(\omega) = v_P(g)$ the order of ω at P. There always exist nonzero differential forms on any compact Riemann-surface, see Springer [8].

We form the free abelian group generated by the points of Π_0, we call it the group of divisors of Π_0. A divisor θ is then a formal sum

$$\theta = \sum_{P \in \Pi_0} \lambda_P \mathrm{P}$$

with $\lambda_P \in \mathbb{Z}$, only finitely many of them being nonzero. The order of the divisor θ is defined to be

$$|\,\theta\,| = \sum_{P \in \Pi_0} \lambda_P,$$

which gives us a homomorphism from the group of divisors to the group of integers.

A meromorphic function f or a meromorphic differential form ω may only have finitely many zeros or poles by compactness, so we can define the divisor of f and ω by

$$\theta\,(f) = \sum_{P \in \Pi_0} v_P(f)\mathrm{P}\,, \quad \theta\,(\omega) = \sum_{P \in \Pi_0} v_P(\omega)\mathrm{P}.$$

By the residue theorem for compact Riemann surfaces, we have $|\,\theta\,(f)\,| = 0$, whereas $|\,\theta\,(\omega)\,| = 2(p-1)$ where p is the genus of Π_0 (this is an easy consequence of the Riemann-Roch Theorem, see below).

We have a natural partial order on the group of divisors: $\theta \geq \theta'$ iff $\lambda_P \geq \lambda'_P$ for each P, the order being compatible with the group structure. If $f \in \mathcal{M}$ is a meromorphic function with $\theta(f) \geq 0$, then f is holomorphic, hence constant, since Π_0 is compact.

Finally, for any divisor θ, consider the complex vector space

$$L(\theta) = \{f \in \mathcal{M} \,:\, \theta(f) + \theta \geq 0\}$$

We have then

Theorem 14. (The Riemann-Roch Theorem) For any differential form ω,

$$\dim L(\theta) = \dim L(\theta(\omega) - \theta) + |\,\theta\,| + 1 - p,$$

where p is the genus of Π_0.

□

5.2 Modular forms for subgroups

Suppose G is a subgroup of the modular group Γ of finite index μ. We are interested in meromorphic functions f on $\mathbf{H}$ which show some invariancy under transformations $g \in$ G. The condition $f(gz) = f(z)$ for all possible g is too restrictive, so we only want $f(gz)$ and $f(z)$ have the same zeros and poles – in that case $f(gz)/f(z) = \lambda_g(z)$ is a holomorphic nonzero function on $\mathbf{H}$. Moreover, we require the consistency condition

$$\lambda_{gg'}(z) = \lambda_g(g'z)\lambda_{g'}(z)$$

for $g, g' \in \mathrm{G}$. From the chain rule for derivatives, these conditions are fulfilled by the functions

$$\lambda_g(z) = \mathrm{J}_g(z)^{-k} = \left(\frac{dg}{dz}\right)^{-k} = (cz+d)^{2k} \text{ if } g : z \to \frac{az+b}{cz+d}.$$

Hence we arrive at the same definition as in 3.1: f is called *weakly modular* of weight $2k$ for G, if f is meromorphic on $\mathbf{H}$, and

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^{2k} f(z) \text{ for all } z \to \frac{az+b}{cz+d} \text{ belonging to G.}$$

As before, we want our functions f to be meromorphic also at cusps. The process to check this at infinity was described in Chapter 3.1, and we do not repeat it here. But in this more general setting, the space $\mathbf{H}/\mathrm{G}$ may have finite parabolic vertices as well – suppose $a/c \in \mathbb{Q}$ is such a vertex. We can then find a transformation $g \in \Gamma$ sending a/c to ∞. Now we can check, whether $f \circ g^{-1}$ is meromorphic or holomorphic at the cusp ∞ of the domain $\mathbf{H}/g\mathrm{G}g^{-1}$. If this is the case, we say that f is meromorphic, resp. holomorphic at the cusp a/c. A weakly modular function, which is meromorphic at cusps, is called a modular function of weight $2k$ for G, whereas an everywhere holomorphic modular function is called a modular form of the same weight for G.

5.3 The dimension of the space of forms for a subgroup G

The set of modular forms of some weight $2k$ for a subgroup G of Γ clearly forms a $\mathbb{C}$-vector space, which we shall denote by $M_k(\mathrm{G})$. It is fairly easy to get a sensible upper bound on the dimension of this space, which we present in the next Theorem.

Theorem 15. (i) If $f \in M_k(\mathrm{G})$ then f has $\frac{k\mu}{6}$ zeros (measured in local variables) in the fundamental domain.

(ii) We have

$$\dim M_k(\mathrm{G}) \leq 1 + \left[\frac{k\mu}{6}\right].$$

(Here and in what follows, [ ] will denote the usual integer part function.)

Proof: $g = f^{12}/\Delta^{2k}$ is a meromorphic function on $\mathbf{H}$ and at the cusps, which is invariant under G, hence it is a modular function of weight 0 and can be considered as a function on the surface Π_{G}. By the discussion above, it has as many zeros as poles, counted according to local multiplicities. But by the properties of Δ, g has $2k\mu$ poles in the fundamental domain (remember the fundamental domain consists of μ copies of $\mathbf{D}$, each of them containing one parabolic vertex), so it has also $2k\mu$ zeros, hence (i).

(Notice that applying this above argument to $\mathrm{G} = \Gamma$ gives our basic equation (10): the left hand side counts the number of zeros according to the local multiplicity, whereas the right hand side is $\frac{k\mu}{6} = \frac{k}{6}$. The only problem with this reasoning is that the properties of Δ used above were deduced from the same equation 10. However, we can resolve this circular argument by noticing that in the previous Chapter we proved in a different way that Δ is nonzero on $\mathbf{H}$; the fact, that it has a simple zero at infinity, is a one-line calculation using the Fourier expansions of G_2 and G_3.)

Now suppose that $f_1, \ldots, f_n$ are linearly independent functions in $M_k(\mathrm{G})$. By linear independence, the determinant of the $n \times n$ Wronskian matrix $W_{ij}(z) = f_i^{(j-1)}(z)$ is not identically zero on $\mathbf{H}$, so we can find a suitable regular point $z_0 \in \mathbf{H}$ and a suitable linear combination $f = \sum \alpha_i f_i$ with

$$\begin{aligned} f^{(j)}(z_0) &= 0 \text{ for } 0 \leq j \leq n-2, \\ f^{(n-1)}(z_0) &= 1. \end{aligned}$$

This means that the nonvanishing function $f \in M_k(\mathrm{G})$ has a zero of order $n-1$ at z_0. By (i) this gives $n - 1 \leq \frac{k\mu}{6}$, which proves (ii).

□

However, to get the exact dimension of $M_k(\mathrm{G})$ needs a more involved argument and the 'heavy guns' introduced in 5.1. To fix some notation, suppose that the fundamental domain for G has n elliptic points $P_1, \ldots, P_n$ with orders $e_1, \ldots, e_n$ respectively, and s parabolic vertices $Q_1, \ldots, Q_s$.

If f^* is any meromorphic function on Π_{G}, then it defines a meromorphic function f on $\mathbf{H}$ by $f(P) = f^*(P')$, P being the point on the surface corresponding to P'. We also have two different kinds of order: $v_{P'}(f^*)$ on the surface, and $n_P(f)$ on $\mathbf{H}$, which is defined accordingly. Similarly a differential form ω^* on Π_{G} induces a differential form ω on $\mathbf{H}$, with two kinds of order again. Our first task is to compare these orders with each other.

- At a regular point $P \in \mathbf{H}$, we have a neighbourhood of P, which is mapped conformally onto a neighbourhood of P', so in this case the two types of order for functions and differentials coincide.

- At an elliptic point P with order e, the local parameter $\tilde{z}$ was obtained by first mapping a neighbourhood of P conformally onto a neighbourhood of 0, and then taking $\tilde{z} = z^e$. Let

  $$f^*(\tilde{z}) = \tilde{z}^v \sum_{i=0}^{\infty} a_i \tilde{z}^i$$

  with $v = v_{P'}(f^*)$ and $a_0 \neq 0$. Then

  $$f(z) = z^{n_P(f)} \sum_{i=0}^{\infty} b_i z^i$$

$$= \quad \tilde{f}(z^e) = t^{ev} \sum_{i=0}^{\infty} a_i z^{ei},$$

so $n_P(f) = ev_{P'}(\tilde{f})$. For differential forms we get

$$\begin{aligned} \omega^*(\tilde{z}) &= f(z)dz \\ &= f^*(\tilde{z})\, d\tilde{z} = f^*(z^e) e z^{e-1} dz, \end{aligned}$$

i.e. $n_P(\omega) = ev_{P'}(\omega^*) + e - 1$.

- If Q is a parabolic point, then a very similar calculation shows that $n_Q(f) = v_{Q'}(f^*)$, whereas $n_Q(\omega) = v_{Q'}(\omega^*) + 1$.

Now select an arbitrary, but from now on fixed, nonzero differential form $\omega^* = h^*(z)dz$ on Π_{G}. This induces a G-invariant form $\omega = h(z)dz$ on $\mathbf{H}$, i.e.

$$h(z)dz = h(gz)d(gz) = h(gz)\mathrm{J}_g(z)dz,$$

which gives $h(gz) = (\mathrm{J}_g(z))^{-1}h(z)$. If f is a modular function of weight $2k$, then $g(z) = \frac{f(z)}{(h(z))^k}$ is a G-invariant meromorphic function inducing a meromorphic function g^* on Π_{G}. Now we have a map $\phi : f \to g^*$ from the space of modular functions of weight $2k$ to the space of meromorphic functions on the surface, which is easily seen to be an isomorphism. Hence the dimension of $M_k(\mathrm{G})$ equals the dimension of those meromorphic functions on Π_{G}, for which $(h(z))^k g(z)$ is holomorphic on $\mathbf{H}$ and at the parabolic vertices. In terms of orders this condition reads $n_P(g) + kn_P(h) \geq 0$. Rewriting this for our surface, we get the following conditions:

- At a regular point

$$v_P(g^*) + kv_P(\omega^*) \geq 0.$$

- At an elliptic point of order e

$$v_P(g^*) + kv_P(\omega^*) + \left[k\left(1 - \frac{1}{e}\right)\right] \geq 0.$$

- At a parabolic point

$$v_P(g^*) + kv_P(\omega^*) + k \geq 0.$$

If we now define the divisor

$$\theta_0 = k\theta(\omega^*) + \sum_{i=1}^{n} \left[k\left(1 - \frac{1}{e}\right)\right] P_i + \sum_{i=1}^{s} k\, Q_i,$$

then the conditions above can be summarized as $\theta(g^*) + \theta_0 \geq 0$. So using our earlier notation, $\dim M_k(\mathrm{G}) = \dim L(\theta_0)$, and this formulation shows that the Riemann-Roch Theorem can be applied immediatelly. Indeed, after a short calculation, which we omit, we get our final result:

Theorem 16.

$$\dim M_k(\mathrm{G}) = \begin{cases} 0 & \text{if } k < 0, \\ 1 & \text{if } k = 0, \\ (2k-1)(p-1) + sk + \sum_{i=1}^{n} \left[k\left(1 - \frac{1}{e_i}\right) \right] & \text{if } k > 0, \end{cases}$$

where p is the genus of Π_{G}.

□

For Γ itself (remember $p = 0$ in this case), we get

$$\dim M_k = 1 - k + \left[\frac{k}{2}\right] + \left[\frac{2k}{3}\right],$$

agreeing with our previous result (11).

6 Some applications

6.1 Arithmetic Identities

The most immediate application of modular forms is that of proving certain identities concerning arithmetic functions. Actually, similar questions led Ramanujan to the study of functions which we call modular forms – see his paper [4]. Recall that $\sigma_k(n)$ denotes the sum of kth powers of positive divisors of n, B_k is the kth Bernoulli number.

Theorem 17. For any positive integer n,

(i) $\sigma_7(n) = \sigma_3(n) + 120 \sum_{i=1}^{n-1} \sigma_3(i)\sigma_3(n-i)$,

(ii) $11\sigma_9(n) = 21\sigma_5(n) - 10\sigma_3(n) + 5040 \sum_{i=1}^{n-1} \sigma_3(i)\sigma_5(n-i)$,

(iii) $\sigma_{13}(n) = 11\sigma_9(n) - 10\sigma_3(n) + 2640 \sum_{i=1}^{n-1} \sigma_3(i)\sigma_9(n-i)$,

(iv) $\sigma_{13}(n) = 21\sigma_5(n) - 20\sigma_7(n) + 10080 \sum_{i=1}^{n-1} \sigma_5(i)\sigma_7(n-i)$.

Proof: We have seen in 3.4 that the dimension of the space of modular forms of weight 8, 10, 14 is 1. This gives

$$E_2^2 = E_4,\ E_2E_3 = E_5,\ E_2E_5 = E_7,\ E_3E_4 = E_7,$$

since the constant term at ∞ on each side is 1. Using the expansions given in (16) and comparing coefficients, we get the desired results.

□

In the general case, we only have asymptotic formulae. A typical example is given in

Theorem 18. For any positive integers $k, l > 1$ and n

$$\sigma_{2(k+l)-1}(n) = \frac{4kl}{k+l}\frac{B_{k+l}}{B_k B_l}\sum_{i=0}^{n}\sigma_{2k-1}(i)\sigma_{2l-1}(n-i) + O(n^{k+l}),$$

where we define $\sigma_{2k-1}(0) = (-1)^k \frac{B_k}{4k}$. (Note that the main terms are of order $n^{2(k+l)-1}$.)

Proof:

$E_k E_l$ and E_{k+l} are both modular forms of weight $2(k+l)$ for Γ with constant term 1 at ∞. Hence

$$\phi_{k+l}(z) = E_k(z)E_l(z) - E_{k+l}(z)$$

is a cusp form. If we compare coefficients and use Theorem 11.(i), we obtain the result.

□

6.2 Theta Series

In this final section we cannot give any proper proofs at all – to build up the theory of theta series would require another whole essay. We only give some indication, how to use the theory of modular forms to study quadratic forms. Proofs of the statements can be found in our usual refence Gunning [1] in Chapter VI and in the references given there.

Suppose $A = (a_{ij})$ is an $r \times r$ real symmetric matrix. We associate to it the quadratic form

$$A[X] = X^t A X = \sum_{ij} a_{ij} X_i X_j$$

where X denotes the column vector $(X_1, \ldots, X_r)^t$. The form is said to be *positive definite* if, whenever $X \neq 0$, we have $A[X] > 0$. Diagonalizing A shows, that it is positive definite if and only if there exists a positive constant c with

$$A[X] > c\sum_i X_i^2. \tag{21}$$

Since we are interested in quadratic forms with integer coefficients, we restrict A to be *semi-integral*, meaning that a_{ii} and $2a_{ij}$ are integers. So henceforth we assume A to be symmetric, positive definite and semi-integral.

To such a matrix A we associate its theta series

$$\theta_A(z) = \sum_N e^{\pi i A[N] z}, \tag{22}$$

where N runs through all integral vectors $(n_1, \ldots, n_r)^t$.

Since A is positive definite, an easy argument using (21) shows that θ_A is holomorphic on the upper half plane $\mathbf{H}$.

Let now $\rho(m, A)$ denote the number of distinct integral vectors N with $A[N] = m$. Then

$$\theta_A(z) = \sum_{m=0}^{\infty} \rho(m, A) e^{\pi i m z}. \tag{23}$$

We clearly have $\theta_A(z+2) = \theta_A(z)$, hence to show that θ_A is a modular form for some group G, one must study the relation between $\theta_A\left(-\frac{1}{z}\right)$ and $\theta_A(z)$. This can be done using the Poisson Summation Formula from complex analysis, and the final result is the so-called *Generalized Jacobi Inversion Formula*:

$$\theta_A(z) = \frac{1}{\sqrt{(-iz)^r \det A}} \theta_{A^{-1}}\left(-\frac{1}{z}\right). \tag{24}$$

One can now continue with this expression, and it is possible to obtain good asymptotic results for $\rho(m, A)$ in the general case. We only consider the case $r = 4k$ and $A = I_{4k}$, then $\rho(m, I_{4k})$ represents the number of ways m can be written as a sum of $4k$ squares. Deleting the subscript A, we have by the previous formulae

$$\theta(z+2) = \theta(z), \tag{25}$$

$$\theta\left(-\frac{1}{z}\right) = (-1)^k z^{2k} \theta(z), \tag{26}$$

from which it can be proved, that $\theta(z)$ is a weakly modular form of weight $2k$ for the full congruence subgroup Γ_2. One can also check the behaviour of $\theta(z)$ at the cusps of the corresponding surface, and the result is that $\theta(z)$ represents a modular form of weight $2k$ for Γ_2. Finally using methods similar to the previous section, one arrives at the asymptotic formula

$$\rho(m, I_{4k}) = \frac{4k}{(2^{2k}-1)B_k} \sum_{\substack{d \mid m \\ 2 \mid m-d}} d^{2k-1} + (-1)^k \sum_{2d \mid m} (-1)^d d^{2k-1} + O(m^k).$$

References

[1] Gunning, R. C. : *Lectures on Modular Forms*, Princeton University Press, Princeton, 1962.

[2] Husemöller, D. : *Elliptic Curves*, Graduate Texts in Mathematics 111, Springer, New York, 1987.

[3] Ogg, A. : *Survey of Modular Functions of One Variable* in: *Modular Forms of One Variable I*, Lecture Notes in Mathematics 320, Springer, Berlin, 1973.

[4] Ramanujan, S. : *On Certain Arithmetical Functions*, Trans. Cambridge Phil. Soc. XXII, 1916.

[5] Sarnak, P. : *Some Applications of Modular Forms*, Cambridge Tracts in Mathematics, CUP, 1990.

[6] Serre, J.-P. : *A Course in Arithmetric*, Graduate Texts in Mathematics 7, Springer, New York, 1973.

[7] Shimura, G. : *Introduction to the Arithmetic Theory of Automorphic Functions*, Publications of the Mathematical Society of Japan 11, Iwanami Shoten Publishers and Princeton University Press, 1971.

[8] Springer, G. : *Introduction to Riemann Surfaces*, Addison-Wesley, Reading, Massachusetts, 1956.