An Introduction to Modular Forms

Henri Cohen

Abstract In this course, we introduce the main notions relative to the classical theory of modular forms. A complete treatise in a similar style can be found in the author's book joint with Strömberg (Cohen and Strömberg, Modular Forms: A Classical Approach, Graduate Studies in Math. 179, American Math. Soc. (2017) [1]).

1 Functional Equations

Let f be a complex function defined over some subset D of $\mathbb{C}$. A *functional equation* is some type of equation relating the value of f at any point $z \in D$ to some other point, for instance, $f(z+1) = f(z)$. If γ is some function from D to itself, one can ask more generally that $f(\gamma(z)) = f(z)$ for all $z \in D$ (or even $f(\gamma(z)) = v(\gamma, z)f(z)$ for some known function v). It is clear that $f(\gamma^m(z)) = f(z)$ for all $m \geq 0$, and even for all $m \in \mathbb{Z}$ if γ is invertible, and more generally the set of bijective functions u such that $f(u(z)) = f(z)$ forms a *group*.

Thus, the basic setting of functional equations (at least of the type that we consider) is that we have a group of transformations G of D, that we ask that $f(u(z)) = f(z)$ (or more generally $f(u(z)) = j(u,z)f(z)$ for some known j) for all $u \in G$ and $z \in D$, and we ask for some type of regularity condition on f such as continuity, meromorphy, or holomorphy.

Note that there is a trivial but essential way to construct from scratch functions f satisfying a functional equation of the above type: simply choose any function g and set $f(z) = \sum_{v \in G} g(v(z))$. Since G is a group, it is clear that *formally* $f(u(z)) = f(z)$ for $u \in G$. Of course there are convergence questions to be dealt with, but this is a fundamental construction, which we call *averaging* over the group.

We consider a few fundamental examples.

H. Cohen (✉)
Institut de Mathématiques de Bordeaux, Université de Bordeaux,
351 Cours de la Libération, 33405 Talence Cedex, France
e-mail: Henri.Cohen@math.u-bordeaux.fr

I. Inam and E. Büyükasik (eds.), *Notes from the International Autumn School on Computational Number Theory*, Tutorials, Schools, and Workshops in the Mathematical Sciences, https://doi.org/10.1007/978-3-030-12558-5_1

1.1 Fourier Series

We choose $D = \mathbb{R}$ and $G = \mathbb{Z}$ acting on $\mathbb{R}$ by translations. Thus, we ask that $f(x+1) = f(x)$ for all $x \in \mathbb{R}$. It is well known that this leads to the theory of *Fourier series*: if f satisfies suitable regularity conditions (we need not specify them here since in the context of modular forms they will be satisfied) then f has an expansion of the type

$$f(x) = \sum_{n\in\mathbb{Z}} a(n)e^{2\pi inx}\ ,$$

absolutely convergent for all $x \in \mathbb{R}$, where the *Fourier coefficients* $a(n)$ are given by the formula

$$a(n) = \int_0^1 e^{-2\pi inx} f(x)\,dx\ ,$$

which follows immediately from the orthonormality of the functions $e^{2\pi imx}$ (you may of course replace the integral from 0 to 1 by an integral from z to $z+1$ for any $z \in \mathbb{R}$).

An important consequence of this, easily proved, is the *Poisson summation formula*: define the *Fourier transform* of f by

$$\widehat{f}(x) = \int_{-\infty}^{\infty} e^{-2\pi ixt} f(t)\,dt\ .$$

We ignore all convergence questions, although of course they must be taken into account in any computation.

Consider the function $g(x) = \sum_{n\in\mathbb{Z}} f(x+n)$, which is exactly the averaging procedure mentioned above. Thus $g(x+1) = g(x)$, so g has a Fourier series, and an easy computation shows the following (again omitting any convergence or regularity assumptions):

Proposition 1.1 (Poisson summation) *We have*

$$\sum_{n\in\mathbb{Z}} f(x+n) = \sum_{m\in\mathbb{Z}} \widehat{f}(m)e^{2\pi imx}\ .$$

In particular,

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{m\in\mathbb{Z}} \widehat{f}(m)\ .$$

A typical application of this formula is to the ordinary Jacobi *theta function*: it is well known (prove it otherwise) that the function $e^{-\pi x^2}$ is invariant under Fourier transform. This implies the following:

Proposition 1.2 *If $f(x) = e^{-a\pi x^2}$ for some $a > 0$, then $\widehat{f}(x) = a^{-1/2}e^{-\pi x^2/a}$.*

Proof Simple change of variable in the integral. □

Corollary 1.3 *Define*

$$T(a) = \sum_{n\in\mathbb{Z}} e^{-a\pi n^2}\ .$$

We have the functional equation

$$T(1/a) = a^{1/2}T(a)\ .$$

Proof Immediate from the proposition and Poisson summation. □

This is historically the first example of modularity, which we will see in more detail below.

Exercise 1.4 Set $S = \sum_{n\geq 1} e^{-(n/10)^2}$.

1. Compute numerically S to 100 decimal digits, and show that it is apparently equal to $5\sqrt{\pi} - 1/2$.
2. Show that, in fact, S is not exactly equal to $5\sqrt{\pi} - 1/2$, and using the above corollary give a precise estimate for the difference.

Exercise 1.5
1. Show that the function $f(x) = 1/\cosh(\pi x)$ is also invariant under Fourier transform.
2. In a manner similar to the corollary, define

$$T_2(a) = \sum_{n\in\mathbb{Z}} 1/\cosh(\pi n a)\ .$$

   Show that we have the functional equation

$$T_2(1/a) = aT_2(a)\ .$$

3. Show that, in fact, $T_2(a) = T(a)^2$ (this may be more difficult).
4. Do the same exercise as the previous one by noticing that $S = \sum_{n\geq 1} 1/\cosh(n/10)$ is very close to $5\pi - 1/2$.

Above we have mainly considered Fourier series of functions defined on $\mathbb{R}$. We now consider more generally functions f defined on $\mathbb{C}$ or a subset of $\mathbb{C}$. We again assume that $f(z+1) = f(z)$, i.e., that f is periodic of period 1. Thus (modulo regularity), f has a Fourier series, but the Fourier coefficients $a(n)$ now depend on $y = \Im(z)$:

$$f(x+iy) = \sum_{n\in\mathbb{Z}} a(n;y)e^{2\pi inx} \quad \text{with} \quad a(n;y) = \int_0^1 f(x+iy)e^{-2\pi inx}\,dx\ .$$

If we impose no extra condition on f, the *functions* $a(n;y)$ are quite arbitrary. But in almost all of our applications f will be *holomorphic*; this means that $\partial(f)(z)/\partial\overline{z} = 0$, or equivalently that $(\partial/\partial(x) + i\partial/\partial(y))(f) = 0$. Replacing in the Fourier expansion (recall that we do not worry about convergence issues) gives

$$\sum_{n\in\mathbb{Z}}(2\pi i n a(n; y) + i a'(n; y))e^{2\pi i n x} = 0 ,$$

and hence by uniqueness of the expansion, we obtain the differential equation $a'(n; y) = -2\pi n a(n; y)$, so that $a(n; y) = c(n)e^{-2\pi n y}$ for some constant $c(n)$. This allows us to write cleanly the Fourier expansion of a holomorphic function in the form

$$f(z) = \sum_{n\in\mathbb{Z}} c(n)e^{2\pi i n z} .$$

Note that if the function is only *meromorphic*, the region of convergence will be limited by the closest pole. Consider, for instance, the function $f(z) = 1/(e^{2\pi i z} - 1) = e^{-\pi i z}/(2i\sin(\pi z))$. If we set $y = \Im(z)$, we have $|e^{2\pi i z}| = e^{-2\pi y}$, so if $y > 0$, we have the Fourier expansion $f(z) = -\sum_{n\geq 0} e^{2\pi i n z}$, while if $y < 0$ we have the different Fourier expansion $f(z) = \sum_{n\leq -1} e^{2\pi i n z}$.

2 Elliptic Functions

The preceding section was devoted to periodic functions. We now assume that our functions are defined on some subset of $\mathbb{C}$ and assume that they are *doubly periodic*: this can be stated either by saying that there exist two $\mathbb{R}$-linearly independent complex numbers ω_1 and ω_2 such that $f(z + \omega_i) = f(z)$ for all z and $i = 1, 2$, or equivalently by saying that there exists a *rank* 2 *lattice* Λ in $\mathbb{C}$ (here $\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$) such that for any $\lambda \in \Lambda$, we have $f(z + \lambda) = f(z)$.

Note in passing that if $\omega_1/\omega_2 \in \mathbb{Q}$ this is equivalent to (single) periodicity, and if $\omega_1/\omega_2 \in \mathbb{R} \setminus \mathbb{Q}$ the set of periods would be dense so the only “doubly periodic” (at least continuous) functions would essentially reduce to functions of one variable. For a similar reason, there do not exist nonconstant continuous functions which are triply periodic.

In the case of simply periodic functions considered above there already existed some natural functions such as $e^{2\pi i n x}$. In the doubly periodic case no such function exists (at least on an elementary level), so we have to construct them, and for this we use the standard averaging procedure seen and used above. Here, the group is the lattice Λ, so we consider functions of the type $f(z) = \sum_{\omega\in\Lambda}\phi(z + \omega)$. For this to converge $\phi(z)$ must tend to 0 sufficiently fast as $|z|$ tends to infinity, and since this is a double sum (Λ is a two-dimensional lattice), it is easy to see by comparison with an integral (assuming $|\phi(z)|$ is regularly decreasing) that $|\phi(z)|$ should decrease at least like $1/|z|^\alpha$ for $\alpha > 2$. Thus, a first reasonable definition is to set

$$f(z) = \sum_{\omega \in \Lambda} \frac{1}{(z+\omega)^3} = \sum_{(m,n) \in \mathbb{Z}^2} \frac{1}{(z+m\omega_1+n\omega_2)^3}\ .$$

This will indeed be a doubly periodic function, and by normal convergence it is immediate to see that it is a meromorphic function on $\mathbb{C}$ having only poles for $z \in \Lambda$, so this is our first example of an *elliptic function*, which is by definition a doubly periodic function which is meromorphic on $\mathbb{C}$. Note for future reference that since $-\Lambda = \Lambda$ this specific function f is odd: $f(-z) = -f(z)$.

However, this is not quite the basic elliptic function that we need. We can integrate term by term, as long as we choose constants of integration such that the integrated series continues to converge. To avoid stupid multiplicative constants, we integrate $-2f(z)$: all antiderivatives of $-2/(z+\omega)^3$ are of the form $1/(z+\omega)^2 + C(\omega)$ for some constant $C(\omega)$, and hence to preserve convergence we will choose $C(0) = 0$ and $C(\omega) = -1/\omega^2$ for $\omega \neq 0$: indeed, $|1/(z+\omega)^2 - 1/\omega^2|$ is asymptotic to $2|z|/|\omega^3|$ as $|\omega| \to \infty$, so we are again in the domain of normal convergence. We will thus define

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda \setminus \{0\}} \left(\frac{1}{(z+\omega)^2} - \frac{1}{\omega^2} \right)\ ,$$

the *Weierstrass $\wp$-function*.

By construction $\wp'(z) = -2f(z)$, where f is the function constructed above, so $\wp'(z+\omega) = \wp'(z)$ for any $\omega \in \Lambda$, and hence $\wp(z+\omega) = \wp(z) + D(\omega)$ for some constant $D(\omega)$ depending on ω but not on z. Note a slightly subtle point here: we use the fact that $\mathbb{C} \setminus \Lambda$ is *connected*. Do you see why?

Now as before it is clear that $\wp(z)$ is an even function: thus, setting $z = -\omega/2$ we have $\wp(\omega/2) = \wp(-\omega/2) + D(\omega) = \wp(\omega/2) + D(\omega)$, so $D(\omega) = 0$, and hence $\wp(z+\omega) = \wp(z)$ and $\wp$ is indeed an elliptic function. There is a mistake in this reasoning: do you see it?

Since $\wp$ has poles on Λ, we cannot reason as we do when $\omega/2 \in \Lambda$. Fortunately, this does not matter: since $\omega_i/2 \notin \Lambda$ for $i = 1, 2$, we have shown at least that $D(\omega_i) = 0$, and hence that $\wp(z+\omega_i) = \wp(z)$ for $i = 1, 2$, so $\wp$ is doubly periodic (so indeed $D(\omega) = 0$ for *all* $\omega \in \Lambda$).

The theory of elliptic functions is incredibly rich, and whole treatises have been written about them. Since this course is mainly about modular forms, we will simply summarize the main properties, and emphasize those that are relevant to us. All are proved using manipulation of power series and complex analysis, and all the proofs are quite straightforward. For instance,

Proposition 2.1 *Let f be a nonzero elliptic function with period lattice Λ as above, and denote by $P = P_a$ a "fundamental parallelogram" $P_a = \{z = a + x\omega_1 + y\omega_2,\ 0 \le x < 1,\ 0 \le y < 1\}$, where a is chosen so that the boundary of P_a does not contain any zeros or poles of f (see Fig. 1).*

1. *The number of zeros of f in P is equal to the number of poles (counted with multiplicity), and this number is called the* order *of f.*

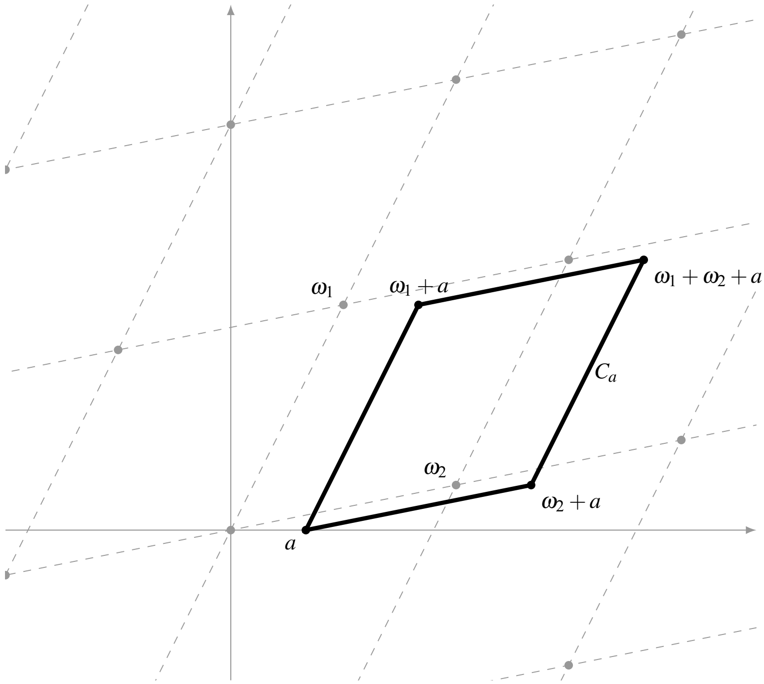

Fig. 1 Fundamental Parallelogram P_a

2. *The sum of the residues of f at the poles in P is equal to* 0.
3. *The sum of the zeros and poles of f in P belongs to Λ.*
4. *If f is nonconstant its order is at least* 2.

Proof For (1), (2), and (3), simply integrate $f(z)$, $f'(z)/f(z)$, and $zf'(z)/f(z)$ along the boundary of P and use the residue theorem. For (4), we first note that by (2) f cannot have order 1 since it would have a simple pole with residue 0. But it also cannot have order 0: this would mean that f has no pole, so it is an entire function, and since it is doubly periodic its values are those taken in the topological closure of P which is compact, so f is *bounded*. By a famous theorem of Liouville (of which this is the no less most famous application), it implies that f is constant, contradicting the assumption of (4). □

Note that clearly $\wp$ has order 2, and the last result shows that we cannot find an elliptic function of order 1. Note however the following:

Exercise 2.2 1. By integrating term by term the series defining $-\wp(z)$ show that if we define the *Weierstrass zeta function*

$$\zeta(z)=\frac{1}{z}+\sum_{\omega\in\Lambda\setminus\{0\}}\left(\frac{1}{z+\omega}-\frac{1}{\omega}+\frac{z}{\omega^2}\right)\ ,$$

this series converges normally on any compact subset of $\mathbb{C} \setminus \Lambda$ and satisfies $\zeta'(z) = -\wp(z)$.

2. Deduce that there exist constants η_1 and η_2 such that $\zeta(z+\omega_1) = \zeta(z) + \eta_1$ and $\zeta(z+\omega_2) = \zeta(z) + \eta_2$, so that if $\omega = m\omega_1 + n\omega_2$, we have $\zeta(z+\omega) = \zeta(z) + m\eta_1 + n\eta_2$. Thus, ζ (which would be of order 1) is not doubly periodic but only quasi-doubly periodic: this is called a *quasi-elliptic function*.
3. By integrating around the usual fundamental parallelogram, show the important relation due to Legendre:

$$\omega_1\eta_2 - \omega_2\eta_1 = \pm 2\pi i \ ,$$

the sign depending on the ordering of ω_1 and ω_2.

The main properties of $\wp$ that we want to mention are as follows: First, for z sufficiently small and $\omega \neq 0$, we can expand

$$\frac{1}{(z+\omega)^2} = \sum_{k\geq 0}(-1)^k(k+1)z^k\frac{1}{\omega^{k+2}} \ ,$$

so

$$\wp(z) = \frac{1}{z^2} + \sum_{k\geq 1}(-1)^k(k+1)z^k G_{k+2}(\Lambda) \ ,$$

where we have set

$$G_k(\Lambda) = \sum_{\omega\in\Lambda\setminus\{0\}}\frac{1}{\omega^k} \ ,$$

which are called *Eisenstein series* of weight k. Since Λ is symmetrical, it is clear that $G_k = 0$ if k is odd, so the expansion of $\wp(z)$ around $z = 0$ is given by

$$\wp(z) = \frac{1}{z^2} + \sum_{k\geq 1}(2k+1)z^{2k}G_{2k+2}(\Lambda) \ .$$

Second, one can show that *all* elliptic functions are simply rational functions in $\wp(z)$ and $\wp'(z)$, so we need not look any further in our construction.

Third, and this is probably one of the most important properties of $\wp(z)$, it satisfies a *differential equation* of order 1: the proof is as follows. Using the above Taylor expansion of $\wp(z)$, it is immediate to check that

$$F(z) = \wp'(z)^2 - (4\wp(z)^3 - g_2(\Lambda)\wp(z) - g_3(\Lambda))$$

has an expansion around $z = 0$ beginning with $F(z) = c_1 z + \cdots$, where we have set $g_2(\Lambda) = 60G_4(\Lambda)$ and $g_3(\Lambda) = 140G_6(\Lambda)$. In addition, F is evidently an elliptic function, and since it has no pole at $z = 0$ it has no poles on Λ, and hence no poles at all, so it has order 0. Thus, by Proposition 2.1 (4) f is constant, and since by construction it vanishes at 0 it is identically 0. Thus, $\wp$ satisfies the differential equation

$$\wp'(z)^2 = 4\wp(z)^3 - g_2(\Lambda)\wp(z) - g_3(\Lambda)\ .$$

A fourth and somewhat surprising property of the function $\wp(z)$ is connected to the theory of *elliptic curves*: the above differential equation shows that $(\wp(z), \wp'(z))$ parametrizes the cubic curve $y^2 = 4x^3 - g_2x - g_3$, which is the general equation of an elliptic curve (you do not need to know the theory of elliptic curves for what follows). Thus, if z_1 and z_2 are in $\mathbb{C} \setminus \Lambda$, the two points $P_i = (\wp(z_i), \wp'(z_i))$ for $i = 1, 2$ are on the curve, and hence if we draw the line through these two points (the tangent to the curve if they are equal), it is immediate to see from Proposition 2.1 (3) that the third point of intersection corresponds to the parameter $-(z_1 + z_2)$, and can of course be computed as a rational function of the coordinates of P_1 and P_2. It follows that $\wp(z)$ (and $\wp'(z)$) possess an *addition formula* expressing $\wp(z_1 + z_2)$ in terms of the $\wp(z_i)$ and $\wp'(z_i)$.

Exercise 2.3 Find this addition formula. You will have to distinguish the cases $z_1 = z_2$, $z_1 = -z_2$, and $z_1 \neq \pm z_2$.

An interesting corollary of the differential equation for $\wp(z)$, which we will prove in a different way below, is a *recursion* for the Eisenstein series $G_{2k}(\Lambda)$:

Proposition 2.4 *We have the recursion for $k \geq 4$:*

$$(k-3)(2k-1)(2k+1)G_{2k} = 3 \sum_{2 \leq j \leq k-2} (2j-1)(2(k-j)-1)G_{2j}G_{2(k-j)}\ .$$

Proof Taking the derivative of the differential equation and dividing by $2\wp'$, we obtain $\wp''(z) = 6\wp(z)^2 - g_2(\Lambda)/2$. If we set by convention $G_0(\Lambda) = -1$ and $G_2(\Lambda) = 0$, and for notational simplicity omit Λ which is fixed, we have $\wp(z) = \sum_{k \geq -1}(2k+1)z^{2k}G_{2k+2}$, so on the one hand

$$\wp''(z) = \sum_{k \geq -1} (2k+1)(2k)(2k-1)z^{2k-2}G_{2k+2}\ ,$$

and on the other hand $\wp(z)^2 = \sum_{K \geq -2} a(K)z^{2K}$ with

$$a(K) = \sum_{k_1+k_2=K} (2k_1+1)(2k_2+1)G_{2k_1+2}G_{2k_2+2}\ .$$

Replacing in the differential equation, it is immediate to check that the coefficients agree up to z^2, and for $K \geq 2$, we have the identification

$$6 \sum_{\substack{k_1+k_2=K \\ k_i \geq -1}} (2k_1+1)(2k_2+1)G_{2k_1+2}G_{2k_2+2} = (2K+3)(2K+2)(2K+1)G_{2K+4}$$

which is easily seen to be equivalent to the recursion of the proposition using $G_0 = -1$ and $G_2 = 0$. $\square$

For instance,

$$G_8 = \frac{3}{7}G_4^2 \quad G_{10} = \frac{5}{11}G_4G_6 \quad G_{12} = \frac{18G_4^3 + 25G_6^2}{143}\ ,$$

and more generally this implies that G_{2k} is a *polynomial* in G_4 and G_6 with rational coefficients which are *independent* of the lattice Λ.

As other corollary, we note that if we choose $\omega_2 = 1$ and $\omega_1 = iT$ with T tending to $+\infty$, then the definition $G_{2k}(\Lambda) = \sum_{(m,n)\in\mathbb{Z}^2\setminus\{(0,0)\}}(m\omega_1 + n\omega_2)^{-2k}$ implies that $G_{2k}(\Lambda)$ will tend to $\sum_{n\in\mathbb{Z}\setminus\{0\}} n^{-2k} = 2\zeta(2k)$, where ζ is the Riemann zeta function. If follows that for all $k \geq 2$, $\zeta(2k)$ is a polynomial in $\zeta(4)$ and $\zeta(6)$ with rational coefficients. Of course this is a weak but nontrivial result, since we know that $\zeta(2k)$ is a rational multiple of π^{2k}.

To finish this section on elliptic functions and make the transition to modular forms, we write explicitly $\Lambda = \Lambda(\omega_1, \omega_2)$ and by abuse of notation $G_{2k}(\omega_1, \omega_2) := G_{2k}(\Lambda(\omega_1, \omega_2))$, and we consider the dependence of G_{2k} on ω_1 and ω_2. We note two evident facts: first, $G_{2k}(\omega_1, \omega_2)$ is *homogeneous* of degree $-2k$: for any nonzero complex number λ, we have $G_{2k}(\lambda\omega_1, \lambda\omega_2) = \lambda^{-2k}G_{2k}(\omega_1, \omega_2)$. In particular, $G_{2k}(\omega_1, \omega_2) = \omega_2^{-2k}G_{2k}(\omega_1/\omega_2, 1)$. Second, a general $\mathbb{Z}$-basis of Λ is given by $(\omega_1', \omega_2') = (a\omega_1 + b\omega_2, c\omega_1 + d\omega_2)$ with a, b, c, and d integers such that $ad - bc = \pm 1$. If we choose an *oriented* basis such that $\Im(\omega_1/\omega_2) > 0$, we, in fact, have $ad - bc = 1$.

Thus, $G_{2k}(a\omega_1 + b\omega_2, c\omega_1 + d\omega_2) = G_{2k}(\omega_1, \omega_2)$, and using homogeneity this can be written as

$$(c\omega_1 + d\omega_2)^{-2k}G_{2k}\left(\frac{a\omega_1 + b\omega_2}{c\omega_1 + d\omega_2}, 1\right) = \omega_2^{-2k}G_{2k}\left(\frac{\omega_1}{\omega_2}, 1\right)\ .$$

Thus, if we set $\tau = \omega_1/\omega_2$ and by an additional abuse of notation abbreviate $G_{2k}(\tau, 1)$ to $G_{2k}(\tau)$, we have by definition

$$G_{2k}(\tau) = \sum_{(m,n)\in\mathbb{Z}^2\setminus\{(0,0)\}} (m\tau + n)^{-2k}\ ,$$

and we have shown the following *modularity* property:

Proposition 2.5 *For any* $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$*, the group of* 2×2 *integer matrices of determinant* 1*, and any* $\tau \in \mathbb{C}$ *with* $\Im(\tau) > 0$*, we have*

$$G_{2k}\left(\frac{a\tau + b}{c\tau + d}\right) = (c\tau + d)^{2k}G_{2k}(\tau)\ .$$

This will be our basic definition of (weak) modularity.

3 Modular Forms and Functions

3.1 Definitions

Let us introduce some notation:

• We denote by Γ the *modular group* $\mathrm{SL}_2(\mathbb{Z})$. Note that properly speaking the modular group should be the group of transformations $\tau \mapsto (a\tau + b)/(c\tau + d)$, which is isomorphic to the quotient of $\mathrm{SL}_2(\mathbb{Z})$ by the equivalence relation saying that M and $-M$ are equivalent, but for this course, we will stick to this definition. If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, we will, of course, write $\gamma(\tau)$ for $(a\tau + b)/(c\tau + d)$.

• The *Poincaré upper half-plane* $\mathscr{H}$ is the set of complex numbers τ such that $\Im(\tau) > 0$. Since for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$, we have $\Im(\gamma(\tau)) = \Im(\tau)/|c\tau + d|^2$, we see that Γ is a group of transformations of $\mathscr{H}$ (more generally so is $\mathrm{SL}_2(\mathbb{R})$, there is nothing special about $\mathbb{Z}$).

• The completed upper half-plane $\overline{\mathscr{H}}$ is by definition $\overline{\mathscr{H}} = \mathscr{H} \cup \mathbb{P}_1(\mathbb{Q}) = \mathscr{H} \cup \mathbb{Q} \cup \{i\infty\}$. Note that this is *not* the closure in the topological sense since we do not include any real irrational numbers.

Definition 3.1 Let $k \in \mathbb{Z}$ and let F be a function from $\mathscr{H}$ to $\mathbb{C}$.

1. We will say that F is *weakly modular* of weight k for Γ if for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ and all $\tau \in \mathscr{H}$, we have
$$F(\gamma(\tau)) = (c\tau + d)^k F(\tau) \ .$$
2. We will say that F is a modular *form* if, in addition, F is holomorphic on $\mathscr{H}$ and if $|F(\tau)|$ remains bounded as $\Im(\tau) \to \infty$.
3. We will say that F is a modular *cusp form* if it is a modular form such that $F(\tau)$ tends to 0 as $\Im(\tau) \to \infty$.

We make a number of immediate but important remarks.

Remarks 3.2 1. The Eisenstein series $G_{2k}(\tau)$ are basic examples of modular forms of weight $2k$, which are not cusp forms since $G_{2k}(\tau)$ tends to $2\zeta(2k) \neq 0$ when $\Im(\tau) \to \infty$.
2. With the present definition, it is clear that there are no nonzero modular forms of *odd weight* k, since if k is odd we have $(-c\tau - d)^k = -(c\tau + d)^k$ and $\gamma(\tau) = (-\gamma)(\tau)$. However, when considering modular forms defined on *subgroups* of Γ there may be modular forms of odd weight, so we keep the above definition.
3. Applying modularity to $\gamma = T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, we see that $F(\tau + 1) = F(\tau)$, and hence F has a Fourier series expansion, and if F is holomorphic, by the remark made above in the section on Fourier series, we have an expansion $F(\tau) = \sum_{n\in\mathbb{Z}} a(n)e^{2\pi i n\tau}$ with $a(n) = e^{2\pi n y} \int_0^1 F(x + iy)e^{-2\pi i n x}\,dx$ for any $y > 0$. Thus, if $|F(x + iy)|$ remains bounded as $y \to \infty$ it follows that as $y \to \infty$, we have

$a(n) \le Be^{2\pi ny}$ for a suitable constant B, so we deduce that $a(n) = 0$ whenever $n < 0$ since $e^{2\pi ny} \to 0$. Thus, if F is a modular *form*, we have $F(\tau) = \sum_{n\ge0} a(n)e^{2\pi in\tau}$, and hence $\lim_{\Im(\tau)\to\infty} F(\tau) = a(0)$, so F is a cusp form if and only if $a(0) = 0$.

Definition 3.3 We will denote by $M_k(\Gamma)$, the vector space of modular forms of weight k on Γ (M for Modular of course), and by $S_k(\Gamma)$ the subspace of cusp forms (S for the German Spitzenform, meaning exactly cusp form).

Notation: for any matrix $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $ad - bc > 0$, we will define the weight k *slash operator* $F|_k\gamma$ by

$$F|_k\gamma(\tau) = (ad - bc)^{k/2}(c\tau + d)^{-k}F(\gamma(\tau)) .$$

The reason for the factor $(ad - bc)^{k/2}$ is that $\lambda\gamma$ has the same action on $\mathscr{H}$ as γ, so this makes the formula homogeneous. For instance, F is weakly modular of weight k if and only if $F|_k\gamma = F$ for all $\gamma \in \Gamma$.

We will also use the universal modular form convention of writing q for $e^{2\pi i\tau}$, so that a Fourier expansion is of the type $F(\tau) = \sum_{n\ge0} a(n)q^n$. We use the additional convention that if α is any complex number, q^α will mean $e^{2\pi i\tau\alpha}$.

Exercise 3.4 Let $F(\tau) = \sum_{n\ge0} a(n)q^n \in M_k(\Gamma)$, and let $\gamma = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ be a matrix in $M_2^+(\mathbb{Z})$, i.e., A, B, C, and D are integers and $\Delta = \det(\gamma) = AD - BC > 0$. Set $g = \gcd(A, C)$, let u and v be such that $uA + vC = g$, set $b = uB + vD$, and finally let $\zeta_\Delta = e^{2\pi i/\Delta}$. Prove the matrix identity

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A/g & -v \\ C/g & u \end{pmatrix}\begin{pmatrix} g & b \\ 0 & \Delta/g \end{pmatrix} ,$$

and deduce that we have the more general Fourier expansion

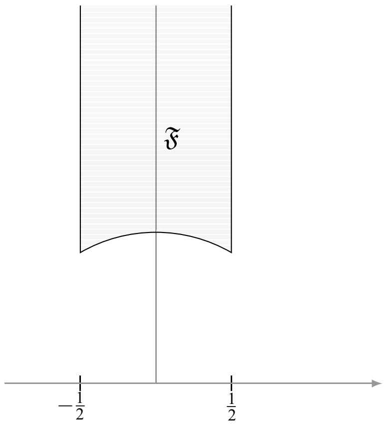

Fig. 2 The fundamental domain, $\mathfrak{F}$, of Γ

$$F|_k\gamma(\tau) = \frac{g^{k/2}}{\Delta^k}\sum_{n\geq 0}\zeta_\Delta^{nbg}a(n)q^{g^2/\Delta}\ ,$$

which is, of course, equal to F if $\Delta = 1$, since then $g = 1$.

3.2 Basic Results

The first fundamental result in the theory of modular forms is that these spaces are *finite dimensional*. The proof uses exactly the same method that we have used to prove the basic results on elliptic functions. We first note that there is a "fundamental domain" (which replaces the fundamental parallelogram, see Fig. 2) for the action of Γ on $\mathscr{H}$, given by

$$\mathfrak{F} = \{\tau \in \mathscr{H},\ -1/2 \leq \Re(\tau) < 1/2,\ |\tau| \geq 1\}\ .$$

The proof that this is a fundamental domain, in other words that any $\tau \in \mathscr{H}$ has a unique image by Γ belonging to $\mathfrak{F}$ is not very difficult and will be omitted. We then integrate $F'(z)/F(z)$ along the boundary of $\mathfrak{F}$, and using modularity, we obtain the following result:

Theorem 3.5 *Let $F \in M_k(\Gamma)$ be a nonzero modular form. For any $\tau_0 \in \mathscr{H}$, denote by $v_{\tau_0}(F)$ the* valuation *of F at τ_0, i.e., the unique integer v such that $F(\tau)/(\tau - \tau_0)^v$ is holomorphic and nonzero at τ_0, and if $F(\tau) = G(e^{2\pi i\tau})$, define $v_{i\infty}(F) = v_0(G)$ (i.e., the number of first vanishing Fourier coefficients of F). We have the formula*

$$v_{i\infty}(F) + \sum_{\tau\in\mathfrak{F}}\frac{v_\tau(F)}{e_\tau} = \frac{k}{12}\ ,$$

where $e_i = 2$, $e_\rho = 3$, and $e_\tau = 1$ otherwise ($\rho = e^{2\pi i/3}$).

This theorem has many important consequences but, as already noted, the most important is that it implies that $M_k(\Gamma)$ is finite dimensional. First, it trivially implies that $k \geq 0$, i.e., there are no modular *forms* of negative weight. In addition, it easily implies the following:

Corollary 3.6 *Let $k \geq 0$ be an even integer. We have*

$$\dim(M_k(\Gamma)) = \begin{cases} \lfloor k/12\rfloor & \textit{if } k \equiv 2 \pmod{12}\ ,\\ \lfloor k/12\rfloor + 1 & \textit{if } k \not\equiv 2 \pmod{12}\ ,\end{cases}$$

$$\dim(S_k(\Gamma)) = \begin{cases} 0 & \textit{if } k < 12\ ,\\ \lfloor k/12\rfloor - 1 & \textit{if } k \geq 12, k \equiv 2 \pmod{12}\ ,\\ \lfloor k/12\rfloor & \textit{if } k \geq 12, k \not\equiv 2 \pmod{12}\ .\end{cases}$$

Since the product of two modular forms is clearly a modular form (of weight the sum of the two weights), it is clear that $M_*(\Gamma) = \bigoplus_k M_k(\Gamma)$ (and similarly $S_*(\Gamma)$) is an algebra, whose structure is easily described as follows:

Corollary 3.7 *We have $M_*(\Gamma) = \mathbb{C}[G_4, G_6]$, and $S_*(\Gamma) = \Delta M_*(\Gamma)$, where Δ is the unique generator of the one-dimensional vector space $S_{12}(\Gamma)$ whose Fourier expansion begins with $\Delta = q + O(q^2)$.*

Thus, for instance, $M_0(\Gamma) = \mathbb{C}$, $M_2(\Gamma) = \{0\}$, $M_4(\Gamma) = \mathbb{C}G_4$, $M_6(\Gamma) = \mathbb{C}G_6$, $M_8(\Gamma) = \mathbb{C}G_8 = \mathbb{C}G_4^2$, $M_{10}(\Gamma) = \mathbb{C}G_{10} = \mathbb{C}G_4G_6$,

$$M_{12}(\Gamma) = \mathbb{C}G_{12} \oplus \mathbb{C}\Delta = \mathbb{C}G_4^3 \oplus \mathbb{C}G_6^2 \ .$$

In particular, we recover the fact proved differently that G_8 is a multiple of G_4^2 (the exact multiple being obtained by computing the Fourier expansions), G_{10} is a multiple of G_4G_6, G_{12} is a linear combination of G_4^3 and G_6^2. Also, we see that Δ is a linear combination of G_4^3 and G_6^2 (we will see this more precisely below).

A basic result on the structure of the modular group Γ is the following:

Proposition 3.8 *Set $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, which acts on $\mathscr{H}$ by the unit translation $\tau \mapsto \tau + 1$, and $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ which acts on $\mathscr{H}$ by the symmetry inversion $\tau \mapsto -1/\tau$. Then, Γ is generated by S and T, with relations generated by $S^2 = -I$ and $(ST)^3 = -I$ (I the identity matrix).*

There are several (easy) proofs of this fundamental result, which we do not give. Simply note that this proposition is essentially equivalent to the fact that the set $\mathfrak{F}$ described above is indeed a fundamental domain.

A consequence of this proposition is that to check whether some function F has the modularity property, it is sufficient to check that $F(\tau + 1) = F(\tau)$ and $F(-1/\tau) = \tau^k F(\tau)$.

Exercise 3.9 (*Bol's identity*). Let F be any continuous function defined on the upper half-plane $\mathscr{H}$, and define $I_0(F, a) = F$ and for any integer $m \geq 1$ and $a \in \overline{\mathscr{H}}$ set:

$$I_m(F, a)(\tau) = \int_a^\tau \frac{(\tau - z)^{m-1}}{(m-1)!} F(z)\,dz \ .$$

1. Show that $I_m(F, a)'(\tau) = I_{m-1}(F, a)(\tau)$, so that $I_m(F, a)$ is an mth antiderivative of F.
2. Let $\gamma \in \Gamma$, and assume that $k \geq 1$ is an integer. Show that

$$I_{k-1}(F, a)|_{2-k}\gamma = I_{k-1}(F|_k\gamma, \gamma^{-1}(a)) \ .$$

3. Deduce that if we set $F_a^* = I_{k-1}(F, a)$, then

$$D^{(k-1)}(F_a^*|_{2-k}\gamma) = F|_k\gamma \ ,$$

where $D = (1/2\pi i)d/d\tau = qd/dq$ is the basic differential operator that we will use (see Sect. 3.10).

4. Assume now that F is weakly modular of weight $k \geq 1$ and holomorphic on $\mathscr{H}$ (in particular, if $F \in M_k(\Gamma)$, but $|F|$ could be unbounded as $\Im(\tau) \to \infty$). Show that

$$(F_a^*|_{2-k}|\gamma)(\tau) = F_a^*(\tau) + P_{k-2}(\tau)\ ,$$

where P_{k-2} is the polynomial of degree less than or equal to $k-2$ given by

$$P_{k-2}(X) = \int_{\gamma^{-1}(a)}^{a} \frac{(X-z)^{k-2}}{(k-2)!} F(z)\,dz\ .$$

What this exercise shows is that the $(k-1)$st derivative of some function which behaves modularly in weight $2-k$ behaves modularly in weight k, and conversely that the $(k-1)$st antiderivative of some function which behaves modularly in weight k behaves modularly in weight k up to addition of a polynomial of degree at most $k-2$. This duality between weights k and $2-k$ is, in fact, a consequence of the *Riemann–Roch theorem*.

Note also that this exercise is the beginning of the fundamental theories of *periods* and of *modular symbols*.

Also, it is not difficult to generalize Bol's identity. For instance, applied to the Eisenstein series G_4 and using Proposition 3.13 below, we obtain:

Proposition 3.10 *1. Set*

$$F_4^*(\tau) = -\frac{\pi^3}{180}\left(\frac{\tau}{i}\right)^3 + \sum_{n\geq 1}\sigma_{-3}(n)q^n\ .$$

We have the functional equation

$$\tau^2 F_4^*(-1/\tau) = F_4^*(\tau) + \frac{\zeta(3)}{2}(1-\tau^2) - \frac{\pi^3}{36}\frac{\tau}{i}\ .$$

2. Equivalently, if we set

$$F_4^{**}(\tau) = -\frac{\pi^3}{180}\left(\frac{\tau}{i}\right)^3 - \frac{\pi^3}{72}\left(\frac{\tau}{i}\right) + \frac{\zeta(3)}{2} + \sum_{n\geq 1}\sigma_{-3}(n)q^n$$

we have the functional equation

$$F_4^{**}(-1/\tau) = \tau^{-2}F_4^{**}(\tau)\ .$$

Note that the appearance of $\zeta(3)$ comes from the fact that, up to a multiplicative constant, the L-function associated to G_4 is equal to $\zeta(s)\zeta(s-3)$, whose value at $s=3$ is equal to $-\zeta(3)/2$.

3.3 The Scalar Product

We begin by the following exercise:

Exercise 3.11 1. Denote by $d\mu = dxdy/y^2$ a measure on $\mathscr{H}$, where as usual x and y are the real and imaginary parts of $\tau \in \mathscr{H}$. Show that this measure is invariant under $\mathrm{SL}_2(\mathbb{R})$.
2. Let f and g be in $M_k(\Gamma)$. Show that the function $F(\tau) = f(\tau)\overline{g(\tau)}y^k$ is invariant under the modular group Γ.

It follows, in particular, from this exercise that if $F(\tau)$ is any integrable function which is invariant by the modular group Γ, the integral $\int_{\Gamma\backslash\mathscr{H}} F(\tau)d\mu$ makes sense if it converges. Since $\mathfrak{F}$ is a fundamental domain for the action of Γ on $\mathscr{H}$, this can also be written as $\int_{\mathfrak{F}} F(\tau)d\mu$. Thus, it follows from the second part that we can define

$$< f, g >= \int_{\Gamma\backslash\mathscr{H}} f(\tau)\overline{g(\tau)}y^k \frac{dxdy}{y^2} ,$$

whenever this converges.

It is immediate to show that a necessary and sufficient condition for convergence is that at least one of f and g be a cusp form, i.e., lies in $S_k(\Gamma)$. In particular, it is clear that this defines a *scalar product* on $S_k(\Gamma)$ called the Petersson scalar product. In addition, any cusp form in $S_k(\Gamma)$ is *orthogonal* to G_k with respect to this scalar product. It is instructive to give a sketch of the simple proof of this fact as given below:

Proposition 3.12 *If $f \in S_k(\Gamma)$, we have $< G_k, f >= 0$.*

Proof Recall that $G_k(\tau) = \sum_{(m,n)\in\mathbb{Z}^2\setminus\{(0,0)\}}(m\tau + n)^{-k}$. We split the sum according to the GCD of m and n: we let $d = \gcd(m, n)$, so that $m = dm_1$ and $n = dn_1$ with $\gcd(m_1, n_1) = 1$. It follows that

$$G_k(\tau) = 2\sum_{d\geq 1} d^{-k}E_k(\tau) = 2\zeta(k)E_k(\tau) ,$$

where $E_k(\tau) = (1/2)\sum_{\gcd(m,n)=1}(m\tau + n)^{-k}$. We thus need to prove that $< E_k, f >= 0$.

On the other hand, denote by Γ_∞ the group generated by T, i.e., translations $\left(\begin{smallmatrix}1 & b\\ 0 & 1\end{smallmatrix}\right)$ for $b \in \mathbb{Z}$. This acts by left multiplication on Γ, and it is immediate to check that a system of representatives for this action is given by matrices $\left(\begin{smallmatrix}u & v\\ m & n\end{smallmatrix}\right)$, where $\gcd(m, n) = 1$ and u and v are chosen arbitrarily (but only once for each pair (m, n)) such that $un - vm = 1$. It follows that, we can write

$$E_k(\tau) = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} (m\tau + n)^{-k} ,$$

where it is understood that $\gamma = \left(\begin{smallmatrix} u & v \\ m & n \end{smallmatrix}\right)$ (the factor $1/2$ has disappeared since γ and $-\gamma$ have the same action on $\mathscr{H}$).

Thus

$$\begin{aligned} < E_k, f > &= \int_{\Gamma\backslash\mathscr{H}} \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} (m\tau+n)^{-k}\overline{f(\tau)}y^k \frac{dxdy}{y^2} \\ &= \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} \int_{\Gamma\backslash\mathscr{H}} (m\tau+n)^{-k}\overline{f(\tau)}y^k \frac{dxdy}{y^2} \ . \end{aligned}$$

Now note that by modularity $f(\tau) = (m\tau+n)^{-k}f(\gamma(\tau))$, and since $\Im(\gamma(\tau)) = \Im(\tau)/|m\tau+n|^2$, it follows that

$$(m\tau+n)^{-k}\overline{f(\tau)}y^k = \overline{f(\gamma(\tau))}\Im(\gamma(\tau))^k \ .$$

Thus, since $d\mu = dxdy/y^2$ is an invariant measure, we have

$$< E_k, f > = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} \int_{\Gamma\backslash\mathscr{H}} \overline{f(\gamma(\tau))}\Im(\gamma(\tau))^k d\mu = \int_{\Gamma_\infty\backslash\mathscr{H}} \overline{f(\tau)}y^k \frac{dxdy}{y^2} \ .$$

Since Γ_∞ is simply the group of integer translations, a fundamental domain for $\Gamma_\infty\backslash\mathscr{H}$ is simply the vertical strip $[0,1]\times[0,\infty[$, so that

$$< E_k, f >= \int_0^\infty y^{k-2}dy \int_0^1 \overline{f(x+iy)}dx \ ,$$

which trivially vanishes since the inner integral is simply the conjugate of the constant term in the Fourier expansion of f, which is 0 since $f \in S_k(\Gamma)$.

The above procedure (replacing the complicated fundamental domain of $\Gamma\backslash\mathscr{H}$ by the trivial one of $\Gamma_\infty\backslash\mathscr{H}$) is very common in the theory of modular forms and is called *unfolding*.

3.4 Fourier Expansions

The Fourier expansions of the Eisenstein series $G_{2k}(\tau)$ are easy to compute. The result is the following:

Proposition 3.13 *For $k \geq 4$ even, we have the Fourier expansion*

$$G_k(\tau) = 2\zeta(k) + 2\frac{(2\pi i)^k}{(k-1)!}\sum_{n\geq 1}\sigma_{k-1}(n)q^n \ ,$$

where $\sigma_{k-1}(n) = \sum_{d\mid n,\ d>0} d^{k-1}$.

Since we know that when k is even $2\zeta(k) = -(2\pi i)^k B_k/k!$, where B_k is the k-th Bernoulli number defined by

$$\frac{t}{e^t - 1} = \sum_{k\geq 0} \frac{B_k}{k!} t^k ,$$

it follows that $G_k = 2\zeta(k)E_k$, with

$$E_k(\tau) = 1 - \frac{2k}{B_k} \sum_{n\geq 1} \sigma_{k-1}(n) q^n .$$

This is the normalization of Eisenstein series that we will use. For instance,

$$\begin{aligned}
E_4(\tau) &= 1 + 240 \sum_{n\geq 1} \sigma_3(n) q^n , \\
E_6(\tau) &= 1 - 504 \sum_{n\geq 1} \sigma_5(n) q^n , \\
E_8(\tau) &= 1 + 480 \sum_{n\geq 1} \sigma_7(n) q^n .
\end{aligned}$$

In particular, the relations given above which follow from the dimension formula become much simpler and are obtained simply by looking at the first terms in the Fourier expansion:

$$E_8 = E_4^2 , \quad E_{10} = E_4 E_6 , \quad E_{12} = \frac{441 E_4^3 + 250 E_6^2}{691} , \quad \Delta = \frac{E_4^3 - E_6^2}{1728} .$$

Note that the relation $E_4^2 = E_8$ (and the others) implies a highly nontrivial relation between the sum of divisors function: if we set by convention $\sigma_3(0) = 1/240$, so that $E_4(\tau) = \sum_{n\geq 0} \sigma_3(n) q^n$, we have

$$E_8(\tau) = E_4^2(\tau) = 240^2 \sum_{n\geq 0} q^n \sum_{0\leq m\leq n} \sigma_3(m)\sigma_3(n-m) ,$$

so that by identification $\sigma_7(n) = 120 \sum_{0\leq m\leq n} \sigma_3(m)\sigma_3(n-m)$, so

$$\sigma_7(n) = \sigma_3(n) + 120 \sum_{1\leq m\leq n-1} \sigma_3(m)\sigma_3(n-m) .$$

It is quite difficult (but not impossible) to prove this directly, i.e., without using at least indirectly the theory of modular forms.

Exercise 3.14 Find a similar relation for $\sigma_9(n)$ using $E_{10} = E_4 E_6$.

This type of reasoning is one of the reasons for which the theory of modular forms is so important (and lots of fun!): if you have a modular form F, you can usually express it in terms of a completely explicit basis of the space to which it belongs since spaces of modular forms are *finite dimensional* (in the present example, the space is one dimensional), and deduce highly nontrivial relations for the Fourier coefficients. We will see a further example of this below for the number $r_k(n)$ of representations of an integer n as a sum of k squares.

Exercise 3.15 1. Prove that for any $k \in \mathbb{C}$, we have the identity

$$\sum_{n\geq 1} \sigma_k(n)q^n = \sum_{n\geq 1} \frac{n^k q^n}{1-q^n}\;,$$

the right-hand side being called a *Lambert series*.

2. Set $F(k) = \sum_{n\geq 1} n^k/(e^{2\pi n}-1)$. Using the Fourier expansions given above, compute explicitly $F(5)$ and $F(9)$.
3. Using Proposition 3.10, compute explicitly $F(-3)$.
4. Using Proposition 3.23 below, compute explicitly $F(1)$.

Note that in this exercise, we only compute $F(k)$ for $k \equiv 1 \pmod 4$. It is also possible but more difficult to compute $F(k)$ for $k \equiv 3 \pmod 4$. For instance, we have

$$F(3) = \frac{\Gamma(1/4)^8}{80(2\pi)^6} - \frac{1}{240}\;.$$

3.5 Obtaining Modular Forms by Averaging

We have mentioned at the beginning of this course that one of the ways to obtain functions satisfying functional equations is to use *averaging* over a suitable group or set: we have seen this for periodic functions in the form of the Poisson summation formula, and for doubly periodic functions in the construction of the Weierstrass $\wp$-function. We can do the same for modular forms, but we must be careful in two different ways. First, we do not want *invariance* by Γ, but we want an automorphy factor $(c\tau+d)^k$. This is easily dealt with by noting that $(d/d\tau)(\gamma(\tau)) = (c\tau+d)^{-2}$: indeed, if ϕ is some function on $\mathscr{H}$, we can define

$$F(\tau) = \sum_{\gamma\in\Gamma} \phi(\gamma(\tau))((d/d\tau)(\gamma(\tau)))^{k/2}\;.$$

Exercise 3.16 Ignoring all convergence questions, by using the chain rule $(f \circ g)' = (f' \circ g)g'$ show that for all $\delta = \left(\begin{smallmatrix} A & B \\ C & D\end{smallmatrix}\right) \in \Gamma$, we have

$$F(\delta(\tau)) = (C\tau + D)^k F(\tau)\;.$$

But the second important way in which we must be careful is that the above construction rarely converges. There are, however, examples where it does converge:

Exercise 3.17 Let $\phi(\tau) = \tau^{-m}$, so that

$$F(\tau) = \sum_{\gamma=\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\in\Gamma} \frac{1}{(a\tau+b)^m(c\tau+d)^{k-m}} .$$

Show that if $2 \le m \le k-2$ and $m \ne k/2$, this series converges normally on any compact subset of $\mathscr{H}$ (i.e., it is majorized by a convergent series with positive terms), so defines a modular form in $M_k(\Gamma)$.

Note that the series converges also for $m = k/2$, but this is more difficult.

One of the essential reasons for non-convergence of the function F is the trivial observation that for a given pair of coprime integers (c, d) there are infinitely many elements $\gamma \in \Gamma$ having (c, d) as their second row. Thus, in general, it seems more reasonable to define

$$F(\tau) = \sum_{\gcd(c,d)=1} \phi(\gamma_{c,d}(\tau))(c\tau+d)^{-k} ,$$

where $\gamma_{c,d}$ is *any fixed* matrix in Γ with second row equal to (c, d). However, we need this to make sense: if $\gamma_{c,d} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ is one such matrix, it is clear that the general matrix having second row equal to (c, d) is $T^n \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} a+nc & b+nd \\ c & d \end{smallmatrix}\right)$, and as usual $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ is translation by 1: $\tau \mapsto \tau + 1$. Thus, an essential necessary condition for our series to make any kind of sense is that the function ϕ be *periodic* of period 1.

The simplest such function is, of course, the constant function 1.

Exercise 3.18 (*See the proof of Proposition* 3.12.) Show that

$$F(\tau) = \sum_{\gcd(c,d)=1} (c\tau+d)^{-k} = 2E_k(\tau) ,$$

where E_k is the normalized Eisenstein series defined above.

But by the theory of Fourier series, we know that periodic functions of period 1 are (infinite) linear combinations of the functions $e^{2\pi i n\tau}$. This leads to the definition of *Poincaré series*:

$$P_k(n;\tau) = \frac{1}{2} \sum_{\gcd(c,d)=1} \frac{e^{2\pi i n\gamma_{c,d}(\tau)}}{(c\tau+d)^k} ,$$

where we note that we can choose any matrix $\gamma_{c,d}$ with bottom row (c, d) since the function $e^{2\pi i n\tau}$ is 1-periodic, so that $P_k(n;\tau) \in M_k(\Gamma)$.

Exercise 3.19 Assume that $k \ge 4$ is even.

1. Show that if $n < 0$, the series defining P_k diverges (wildly in fact).
2. Note that $P_k(0; \tau) = E_k(\tau)$, so that $\lim_{\tau \to i\infty} P_k(0; \tau) = 1$. Show that if $n > 0$, the series converges normally and that we have $\lim_{\tau \to i\infty} P_k(n; \tau) = 0$. Thus, in fact, $P_k(n; \tau) \in S_k(\Gamma)$ if $n > 0$.
3. By using the same *unfolding method* as in Proposition 3.12, show that if $f = \sum_{n \geq 0} a(n)q^n \in M_k(\Gamma)$ and $n > 0$, we have

$$< P_k(n), f >= \frac{(k-2)!}{(4\pi n)^{k-1}} a(n) .$$

It is easy to show that, in fact, the $P_k(n)$ *generate* $S_k(\Gamma)$. We can also compute their *Fourier expansions* as we have done for E_k, but they involve Bessel functions and Kloosterman sums.

3.6 *The Ramanujan Delta Function*

Recall that by definition Δ is the generator of the one-dimensional space $S_{12}(\Gamma)$ whose Fourier coefficient of q^1 is normalized to be equal to 1. By simple computation, we find the first terms in the Fourier expansion of Δ:

$$\Delta(\tau) = q - 24q^2 + 252q^3 - 1472q^4 + \cdots ,$$

with no apparent formula for the coefficients. The nth coefficient is denoted $\tau(n)$ (no confusion with $\tau \in \mathscr{H}$), and called Ramanujan's tau function, and Δ itself is called Ramanujan's Delta function.

Of course, using $\Delta = (E_4^3 - E_6^2)/1728$ and expanding the powers, one can give a complicated but explicit formula for $\tau(n)$ in terms of the functions σ_3 and σ_5, but this is far from being the best way to compute them. In fact, the following exercise already gives a much better method.

Exercise 3.20 Let D be the differential operator $(1/(2\pi i))d/d\tau = qd/dq$.

1. Show that the function $F = 4E_4D(E_6) - 6E_6D(E_4)$ is a modular form of weight 12, then by looking at its constant term show that it is a cusp form, and finally compute the constant c such that $F = c \cdot \Delta$.
2. Deduce the formula

$$\tau(n) = \frac{n}{12}(5\sigma_3(n) + 7\sigma_5(n)) + 70 \sum_{1 \leq m \leq n-1} (2n - 5m)\sigma_3(m)\sigma_5(n-m) .$$

3. Deduce, in particular, the congruences $\tau(n) \equiv n\sigma_5(n) \equiv n\sigma_1(n) \pmod{5}$ and $\tau(n) \equiv n\sigma_3(n) \pmod{7}$.

Although there are much faster methods, this is already a very reasonable way to compute $\tau(n)$.

The cusp form Δ is one of the most important functions in the theory of modular forms. Its first main property, which is not at all apparent from its definition, is that it has a *product expansion*:

Theorem 3.21 *We have*

$$\Delta(\tau) = q\prod_{n\geq 1}(1-q^n)^{24} .$$

Proof We are not going to give a complete proof, but sketch a method which is one of the most natural to obtain the result.

We start backward, from the product $R(\tau)$ on the right-hand side. The logarithm transforms products into sums, but in the case of *functions* f, the *logarithmic derivative* f'/f (more precisely $D(f)/f$, where $D = qd/dq$) also does this, and it is also more convenient. We have

$$D(R)/R = 1 - 24\sum_{n\geq 1}\frac{nq^n}{1-q^n} = 1 - 24\sum_{n\geq 1}\sigma_1(n)q^n$$

as is easily seen by expanding $1/(1-q^n)$ as a geometric series. This is exactly the case $k=2$ of the Eisenstein series E_k, which we have excluded from our discussion for convergence reasons, so we come back to our series G_{2k} (we will divide by the normalizing factor $2\zeta(2) = \pi^2/3$ at the end), and introduce a convergence factor due to Hecke, setting

$$G_{2,s}(\tau) = \sum_{(m,n)\in\mathbb{Z}^2\setminus\{(0,0)\}}(m\tau+n)^{-2}|m\tau+n|^{-2s} .$$

As above this converges for $\Re(s) > 0$, it satisfies

$$G_{2,s}(\gamma(\tau)) = (c\tau+d)^2|c\tau+d|^{2s}G_{2,s}(\tau)$$

and hence, in particular, is periodic of period 1. It is straightforward to compute its Fourier expansion, which we will not do here, and the Fourier expansion shows that $G_{2,s}$ has an *analytic continuation* to the whole complex plane. In particular, the limit as $s \to 0$ makes sense; if we denote it by $G_2^*(\tau)$, by continuity it will of course satisfy $G_2^*(\gamma(\tau)) = (c\tau+d)^2G_2^*(\tau)$, and the analytic continuation of the Fourier expansion that has been computed gives

$$G_2^*(\tau) = \frac{\pi^2}{3}\left(1 - \frac{3}{\pi\Im(\tau)} - 24\sum_{n\geq 1}\sigma_1(n)q^n\right) .$$

Note the essential fact that there is now a *nonanalytic term* $3/(\pi\Im(\tau))$. We will, of course, set the following definition:

Definition 3.22 We define

$$E_2(\tau) = 1 - 24\sum_{n\geq 1}\sigma_1(n)q^n \quad\text{and}\quad E_2^*(\tau) = E_2(\tau) - \frac{3}{\pi\Im(\tau)}\ .$$

Thus, $E_2(\tau) = D(R)/R$, $G_2^*(\tau) = (\pi^2/3)E_2^*(\tau)$, and we have the following:

Proposition 3.23 *For any* $\gamma = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \Gamma$, *we have* $E_2^*(\gamma(\tau)) = (c\tau+d)^2E_2^*(\tau)$. *Equivalently,*

$$E_2(\gamma(\tau)) = (c\tau+d)^2E_2(\tau) + \frac{12}{2\pi i}c(c\tau+d)\ .$$

Proof The first result has been seen above, and the second follows from the formula $\Im(\gamma(\tau)) = \Im(\tau)/|c\tau+d|^2$. □

Exercise 3.24 Show that

$$E_2(\tau) = -24\left(-\frac{1}{24} + \sum_{m\geq 1}\frac{m}{q^{-m}-1}\right)\ .$$

Proof of the theorem. We can now prove the theorem on the product expansion of Δ: noting that $(d/d\tau)\gamma(\tau) = 1/(c\tau+d)^2$, the above formulas imply that if we set $S = R(\gamma(\tau))$, we have

$$\begin{aligned}\frac{D(S)}{S} &= \frac{D(R)}{R}(\gamma(\tau))(d/d\tau)(\gamma(\tau))\\ &= (c\tau+d)^{-2}E_2(\gamma(\tau)) = E_2(\tau) + \frac{12}{2\pi i}\frac{c}{c\tau+d}\\ &= \frac{D(R)}{R}(\tau) + 12\frac{D(c\tau+d)}{c\tau+d}\ .\end{aligned}$$

By integrating and exponentiating, it follows that

$$R(\gamma(\tau)) = (c\tau+d)^{12}R(\tau)\ ,$$

and since clearly R is holomorphic on $\mathscr{H}$ and tends to 0 as $\Im(\tau) \to \infty$ (i.e., as $q \to 0$), it follows that R is a cusp form of weight 12 on Γ, and since $S_{12}(\Gamma)$ is one-dimensional and the coefficient of q^1 in R is 1, we have $R = \Delta$, proving the theorem. □

Exercise 3.25 We have shown in passing that $D(\Delta) = E_2\Delta$. Expanding the Fourier expansion of both sides, show that we have the recursion

$$(n-1)\tau(n) = -24\sum_{1\leq m\leq n-1}\sigma_1(m)\tau(n-m)\ .$$

Exercise 3.26 1. Let $F \in M_k(\Gamma)$, and for some *squarefree* integer N set

$$G(\tau) = \sum_{d \mid N} \mu(d) d^{k/2} F(d\tau) ,$$

where μ is the Möbius function. Show that $G|_k W_N = \mu(N) G$, where $W_N = \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right)$ is the so-called *Fricke involution*.

2. Show that if $N > 1$, the same result is true for $F = E_2$, although E_2 is only quasi-modular.
3. Deduce that if $\mu(N) = (-1)^{k/2-1}$, we have $G(i/\sqrt{N}) = 0$.
4. Applying this to E_2 and using Exercise 3.24, deduce that if $\mu(N) = 1$ and $N > 1$, we have

$$\sum_{\gcd(m,N)=1} \frac{m}{e^{2\pi m/\sqrt{N}} - 1} = \frac{\phi(N)}{24} ,$$

where $\phi(N)$ is Euler's totient function.

5. Using directly the functional equation of E_2^*, show that for $N = 1$ there is an additional term $-1/(8\pi)$, i.e., that

$$\sum_{m \geq 1} \frac{m}{e^{2\pi m} - 1} = \frac{1}{24} - \frac{1}{8\pi} .$$

3.7 *Product Expansions and the Dedekind Eta Function*

We continue our study of product expansions. We first mention an important identity due to Jacobi, the triple product identity, as well as some consequences:

Theorem 3.27 (Triple product identity) *If* $|q| < 1$ *and* $u \neq 0$, *we have*

$$\prod_{n \geq 1} (1 - q^n)(1 - q^n u) \prod_{n \geq 0} (1 - q^n/u) = \sum_{k \geq 0} (-1)^k (u^k - u^{-(k+1)}) q^{k(k+1)/2} .$$

Proof (sketch): Denote by $L(q, u)$ the left-hand side. We have clearly $L(q, u/q) = -uL(q, u)$, and since one can write $L(q, u) = \sum_{k \in \mathbb{Z}} a_k(q) u^k$ this implies the recursion $a_k(q) = -q^k a_{k-1}(q)$, so $a_k(q) = (-1)^k q^{k(k+1)/2} a_0(q)$, and separating $k \geq 0$ and $k < 0$ this shows that

$$L(q, u) = a_0(q) \sum_{k \geq 0} (-1)^k (u^k - u^{-(k+1)}) q^{k(k+1)/2} .$$

The slightly longer part is to show that $a_0(q) = 1$: this is done by setting $u = i/q^{1/2}$ and $u = 1/q^{1/2}$, which after a little computation implies that $a(q^4) = a(q)$, and from there it is immediate to deduce that $a(q)$ is a constant, and equal to 1. □

To give the next corollaries, we need to define the *Dedekind eta function* $\eta(\tau)$, by

$$\eta(\tau) = q^{1/24} \prod_{n\geq 1}(1-q^n)\ ,$$

(recall that $q^\alpha = e^{2\pi i\alpha\tau}$). Thus, by definition $\eta(\tau)^{24} = \Delta(\tau)$. Since $\Delta(-1/\tau) = \tau^{12}\Delta(\tau)$, it follows that $\eta(-1/\tau) = c\cdot(\tau/i)^{1/2}\eta(\tau)$ for some 24th root of unity c (where we always use the principal determination of the square root), and since we see from the infinite product that $\eta(i)\neq 0$, replacing τ by i shows that, in fact, $c=1$. Thus, η satisfies the two basic modular equations

$$\eta(\tau+1) = e^{2\pi i/24}\eta(\tau) \quad\text{and}\quad \eta(-1/\tau) = (\tau/i)^{1/2}\eta(\tau)\ .$$

Of course, we have more generally

$$\eta(\gamma(\tau)) = v_\eta(\gamma)(c\tau+d)^{1/2}\eta(\tau)$$

for any $\gamma\in\Gamma$, with a complicated 24th root of unity $v_\eta(\gamma)$, so η is in some (reasonable) sense a modular form of weight $1/2$, similar to the function θ that we introduced at the very beginning.

The triple product identity immediately implies the following two identities:

Corollary 3.28 *We have*

$$\eta(\tau) = q^{1/24}\left(1+\sum_{k\geq 1}(-1)^k\left(q^{k(3k-1)/2}+q^{k(3k+1)/2}\right)\right) \quad \textit{and}$$
$$\eta(\tau)^3 = q^{1/8}\sum_{k\geq 0}(-1)^k(2k+1)q^{k(k+1)/2}\ .$$

Proof In the triple product identity, replace (u,q) by $(1/q,q^3)$, we obtain

$$\prod_{n\geq 1}(1-q^{3n})(1-q^{3n-1})\prod_{n\geq 0}(1-q^{3n+1}) = \sum_{k\geq 0}(-1)^k(q^{-k}-q^{k+1})q^{3k(k+1)/2}\ .$$

The left-hand side is clearly equal to $\eta(\tau)$, and the right-hand side to

$$\begin{aligned}1-q+&\sum_{k\geq 1}(-1)^k\left(q^{k(3k+1)/2}-q^{(k+1)(3k+2)/2}\right)\\ &= 1+\sum_{k\geq 1}(-1)^k q^{k(3k+1)/2} - q + \sum_{k\geq 2}(-1)^k q^{k(3k-1)/2}\ ,\end{aligned}$$

giving the formula for $\eta(\tau)$. For the second formula, divide the triple product identity by $1-1/u$ and make $u\to 1$. ☐

Thus, the first few terms are

$$\prod_{n\ge1}(1-q^n) = 1 - q - q^2 + q^5 + q^7 - q^{12} - q^{15} + \cdots$$
$$\prod_{n\ge1}(1-q^n)^3 = 1 - 3q + 5q^3 - 7q^6 + 9q^{10} - 11q^{15} + \cdots\ .$$

The first identity was proved by L. Euler.

Exercise 3.29 1. Show that $24\Delta D(\eta) = \eta D(\Delta)$, and using the explicit Fourier expansion of η, deduce the recursion

$$\sum_{k\in\mathbb{Z}}(-1)^k(75k^2+25k+2-2n)\tau\left(n-\frac{k(3k+1)}{2}\right) = 0\ .$$

2. Similarly, from $8\Delta D(\eta^3) = \eta^3 D(\Delta)$ deduce the recursion

$$\sum_{k\in\mathbb{Z}}(-1)^k(2k+1)(9k^2+9k+2-2n)\tau\left(n-\frac{k(k+1)}{2}\right) = 0\ .$$

Exercise 3.30 Define the *q-Pochhammer symbol* $(q)_n$ by $(q)_n = (1-q)(1-q^2)\cdots(1-q^n)$.

1. Set $f(a,q) = \prod_{n\ge1}(1-aq^n)$, and define coefficients $c_n(q)$ by setting $f(a,q) = \sum_{n\ge0} c_n(q)a^n$. Show that $f(a,q) = (1-aq)f(aq,q)$, deduce that $c_n(q)(1-q^n) = -q^n c_{n-1}(q)$ and finally the identity

$$\prod_{n\ge1}(1-aq^n) = \sum_{n\ge0}(-1)^n a^n q^{n(n+1)/2}/(q)_n\ .$$

2. Write in terms of the Dedekind eta function, the identities obtained by specializing to $a=1$, $a=-1$, $a=-1/q$, $a=q^{1/2}$, and $a=-q^{1/2}$.
3. Similarly, prove the identity

$$1/\prod_{n\ge1}(1-aq^n) = \sum_{n\ge0} a^n q^n/(q)_n\ ,$$

and once again write in terms of the Dedekind eta function, the identities obtained by specializing to the same five values of a.
4. By multiplying two of the above identities and using the triple product identity, prove the identity

$$\frac{1}{\prod_{n\ge1}(1-q^n)} = \sum_{n\ge0}\frac{q^{n^2}}{(q)_n^2}\ .$$

Note that this last series is the generating function of the *partition function* $p(n)$, so if one wants to make a table of $p(n)$ up to $n = 10{,}000$, say, using the left-hand side would require 10,000 terms, while using the right-hand side only requires 100.

3.8 Computational Aspects of the Ramanujan τ Function

Since its introduction, the Ramanujan tau function $\tau(n)$ has fascinated number theorists. For instance, there is a conjecture due to D. H. Lehmer that $\tau(n) \neq 0$, and an even stronger conjecture (which would imply the former) that for every prime p we have $p \nmid \tau(p)$ (on probabilistic grounds, the latter conjecture is probably false).

To test these conjectures as well as others, it is an interesting computational challenge to *compute* $\tau(n)$ for large n (because of Ramanujan's first two conjectures, i.e., Mordell's theorem that we will prove in Sect. 4 below, it is sufficient to compute $\tau(p)$ for p *prime*).

We can have two distinct goals. The first is to compute a *table* of $\tau(n)$ for $n \le B$, where B is some (large) bound. The second is to compute *individual values* of $\tau(n)$, equivalently of $\tau(p)$ for p prime.

Consider first the construction of a *table*. The use of the first recursion given in the above exercise needs $O(n^{1/2})$ operations per value of $\tau(n)$, and hence $O(B^{3/2})$ operations in all to have a table for $n \le B$.

However, it is well known that the *Fast Fourier Transform* (FFT) allows one to compute products of power series in essentially linear time. Thus, using Corollary 3.28, we can directly write the power series expansion of η^3, and use the FFT to compute its eighth power $\eta^{24} = \Delta$. This will require $O(B\log(B))$ operations, so it is much faster than the preceding method; it is essentially optimal since one needs $O(B)$ time simply to write the result.

Using large computer resources, especially in memory, it is reasonable to construct a table up to $B = 10^{12}$, but not much more. Thus, the problem of computing *individual* values of $\tau(p)$ is important. We have already seen one such method in Exercise 3.20 above, which gives a method for computing $\tau(n)$ in time $O(n^{1+\varepsilon})$ for any $\varepsilon > 0$.

A deep and important theorem of B. Edixhoven, J.-M. Couveignes, et al., says that it is possible to compute $\tau(p)$ in time *polynomial* in $\log(p)$, and, in particular, in time $O(p^{\varepsilon})$ for any $\varepsilon > 0$. Unfortunately, this algorithm is not at all practical, and at least for now, completely useless for us. The only practical and important application is for the computation of $\tau(p)$ modulo some small prime numbers ℓ (typically $\ell < 50$, so far from being sufficient to apply the Chinese Remainder Theorem).

However, there exists an algorithm which takes time $O(n^{1/2+\varepsilon})$ for any $\varepsilon > 0$, so much better than the one of Exercise 3.20, and which is very practical. It is based on the use of the Eichler–Selberg *trace formula*, together with the computation of *Hurwitz class numbers* $H(N)$ (essentially the class numbers of imaginary quadratic orders counted with suitable multiplicity): if we set $H_3(N) = H(4N) + 2H(N)$ (note that $H(4N)$ can be computed in terms of $H(N)$), then for p prime

$$\tau(p) = 28p^6 - 28p^5 - 90p^4 - 35p^3 - 1$$
$$- 128 \sum_{1 \le t < p^{1/2}} t^6(4t^4 - 9pt^2 + 7p^2)H_3(p - t^2)\ .$$

See [1] Exercise 12.13 of Chap. 12 for details. Using this formula and a cluster, it should be reasonable to compute $\tau(p)$ for p of the order of 10^{16}.

3.9 Modular Functions and Complex Multiplication

Although the terminology is quite unfortunate, we cannot change it. By definition, a modular *function* is a function F from $\mathscr{H}$ to $\mathbb{C}$ which is weakly modular of weight 0 (so that $F(\gamma(\tau)) = F(\tau)$, in other words is *invariant* under Γ, or equivalently defines a function from $\Gamma\backslash\mathscr{H}$ to $\mathbb{C}$), meromorphic, including at ∞. This last statement requires some additional explanation, but in simple terms, this means that the Fourier expansion of F has only finitely many Fourier coefficients for negative powers of q: $F(\tau) = \sum_{n \ge n_0} a(n)q^n$, for some (possibly negative) n_0.

A trivial way to obtain modular functions is simply to take the quotient of two modular forms having the same weight. The most important is the j-function defined by

$$j(\tau) = \frac{E_4^3(\tau)}{\Delta(\tau)}\ ,$$

whose Fourier expansion begins by

$$j(\tau) = \frac{1}{q} + 744 + 196884q + 21493760q^2 + \cdots$$

Indeed, one can easily prove the following theorem:

Theorem 3.31 *Let F be a meromorphic function on $\mathscr{H}$. The following are equivalent:*

1. *F is a modular function.*
2. *F is the quotient of two modular forms of equal weight.*
3. *F is a rational function of j.*

Exercise 3.32 1. Noting that Theorem 3.5 is valid more generally for modular functions (with $v_\tau(f) = -r < 0$ if f has a pole of order r at τ) and using the specific properties of $j(\tau)$, compute $v_\tau(f)$ for the functions $j(\tau)$, $j(\tau) - 1728$, and $D(j)(\tau)$, at the points $\rho = e^{2\pi i/3}$, i, $i\infty$, and τ_0 for τ_0 distinct from these three special points.

2. Set $f = f(a, b, c) = D(j)^a/(j^b(j - 1728)^c)$. Show that f is a modular *form* if and only if $2c \le a$, $3b \le 2a$, and $b + c \ge a$, and give similar conditions for f to be a *cusp form*.

3. Show that $E_4 = f(2,1,1)$, $E_6 = f(3,2,1)$, and $\Delta = f(6,4,3)$, so that, for instance, $D(j) = -E_{14} = -E_4^2 E_6/\Delta$.

An important theory linked to modular functions is the theory of *complex multiplication*, which deserves a course in itself. We simply mention one of the basic results.

We will say that a complex number $\tau \in \mathscr{H}$ is a CM point (CM for Complex Multiplication), if it belongs to an imaginary quadratic field, or equivalently if there exist integers a, b, and c with $a \neq 0$ such that $a\tau^2 + b\tau + c = 0$. The first basic theorem is the following:

Theorem 3.33 *If τ is a CM point, then $j(\tau)$ is an algebraic integer.*

Note that this theorem has two parts: the first and most important part is that $j(\tau)$ is algebraic. This is, in fact, easy to prove. The second part is that it is an algebraic *integer*, and this is more difficult. Since any modular function f is a rational function of j, it follows that if this rational function has algebraic coefficients then $f(\tau)$ will be algebraic (but not necessarily integral). Another immediate consequence is the following:

Corollary 3.34 *Let τ be a CM point and define $\Omega_\tau = \eta(\tau)^2$, where η is as usual the Dedekind eta function. For any modular form f of weight k (in fact, f can also be meromorphic), the number $f(\tau)/\Omega_\tau^k$ is algebraic. In fact, $E_4(\tau)/\Omega_\tau^4$ and $E_6(\tau)/\Omega_\tau^6$ are always algebraic* integers.

But the importance of this theorem lies in algebraic number theory. We give the following theorem without explaining the necessary notions:

Theorem 3.35 *Let τ be a CM point, $D = b^2 - 4ac$ its* discriminant*, where we choose* $\gcd(a,b,c) = 1$ *and $K = \mathbb{Q}(\sqrt{D})$. Then, $K(j(\tau))$ is the maximal abelian unramified extension of K, the* Hilbert class field *of K, and $\mathbb{Q}(j(\tau))$ is the* ring class field *of discriminant D. In particular, the degree of the minimal polynomial of the algebraic integer $j(\tau)$ is equal to the* class number $h(D)$ *of the order of discriminant D, and its algebraic conjugates are given by an explicit formula called the* Shimura reciprocity law.

Examples:

$$j((1+i\sqrt{3})/2) = 0 = 1728 - 3(24)^2$$
$$j(i) = 1728 = 12^3 = 1728 - 4(0)^2$$
$$j((1+i\sqrt{7})/2) = -3375 = (-15)^3 = 1728 - 7(27)^2$$
$$j(i\sqrt{2}) = 8000 = 20^3 = 1728 + 8(28)^2$$
$$j((1+i\sqrt{11})/2) = -32768 = (-32)^3 = 1728 - 11(56)^2$$
$$j((1+i\sqrt{163})/2) = -262537412640768000 = (-640320)^3$$

$$
\begin{aligned}
&= 1728 - 163(40133016)^2 \\
j(i\sqrt{3}) &= 54000 = 2(30)^3 = 1728 + 12(66)^2 \\
j(2i) &= 287496 = (66)^3 = 1728 + 8(189)^2 \\
j((1+3i\sqrt{3})/2) &= -12288000 = -3(160)^3 = 1728 - 3(2024)^2 \\
j((1+i\sqrt{15})/2) &= \frac{-191025 - 85995\sqrt{5}}{2} \\
&= \frac{1-\sqrt{5}}{2}\left(\frac{75+27\sqrt{5}}{2}\right)^3 = 1728 - 3\left(\frac{273+105\sqrt{5}}{2}\right)^2
\end{aligned}
$$

Note that we give the results in the above form since show that the functions $j^{1/3}$ and $(j-1728)^{1/2}$ also have interesting arithmetic properties.

The example with $D = -163$ is particularly spectacular:

Exercise 3.36 Using the above table, show that

$$(e^{\pi\sqrt{163}} - 744)^{1/3} = 640320 - \varepsilon\ ,$$

with $0 < \varepsilon < 10^{-24}$, and more precisely that ε is approximately equal to $65628e^{-(5/3)\pi\sqrt{163}}$ (note that $65628 = 196884/3$).

Exercise 3.37 1. Using once again the example of 163, compute heuristically a few terms of the Fourier expansion of j assuming that it is of the form $1/q + \sum_{n\geq 0} c(n)q^n$ with $c(n)$ reasonably small integers using the following method. Set $q = -e^{-\pi\sqrt{163}}$, and let $J = (-640320)^3$ be the exact value of $j((-1+i\sqrt{163})/2)$. By computing $J - 1/q$, one notices that the result is very close to 744, so we guess that $c(0) = 744$. We then compute $(J - 1/q - c(0))/q$ and note that once again the result is close to an integer, giving $c(1)$, and so on. Go as far as you can with this method.

2. Do the same for 67 instead of 163. You will find the same Fourier coefficients (but you can go less far).
3. On the other hand, do the same for 58, starting with J equal to the integer close to $e^{\pi\sqrt{58}}$. You will find a *different* Fourier expansion: it corresponds, in fact, to another modular function, this time defined on a subgroup of Γ, called a *Hauptmodul*.
4. Try to find other rational numbers D such that $e^{\pi\sqrt{D}}$ is close to an integer, and do the same exercise for them (an example where D is not integral is $89/3$).

3.10 Derivatives of Modular Forms

If we differentiate the modular equation $f((a\tau+b)/(c\tau+d)) = (c\tau+d)^k f(\tau)$ with $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ using the operator $D = (1/(2\pi i))d/d\tau$ (which gives simpler formulas than $d/d\tau$ since $D(q^n) = nq^n$), we easily obtain

$$D(f)\left(\frac{a\tau+b}{c\tau+d}\right) = (c\tau+d)^{k+2}\left(D(f)(\tau) + \frac{k}{2\pi i}\frac{c}{c\tau+d}f(\tau)\right) .$$

Thus, the derivative of a weakly modular form of weight k looks like one of weight $k+2$, except that there is an extra term. This term vanishes if $k=0$, so the derivative of a modular function of weight 0 is indeed modular of weight 2 (we have seen above the example of $j(\tau)$ which satisfies $D(j) = -E_{14}/\Delta$).

If $k > 0$ and we really want a true weakly modular form of weight $k+2$ there are two ways to do this. The first one is called the *Serre derivative*:

Exercise 3.38 Using Proposition 3.23, show that if f is weakly modular of weight k, then $D(f) - (k/12)E_2 f$ is weakly modular of weight $k+2$. In particular, if $f \in M_k(\Gamma)$, then $SD_k(f) := D(f) - (k/12)E_2 f \in M_{k+2}(\Gamma)$.

The second method is to set $D^*(f) := D(f) - (k/(4\pi\Im(\tau)))f$ since by Proposition 3.23, we have $D^*(f) = SD_k(f) - (k/12)E_2^* f$. This loses holomorphy, but is very useful in certain contexts.

Note that if more than one modular form is involved, there are more ways to make new modular forms using derivatives.

Exercise 3.39 1. For $i = 1, 2$ let $f_i \in M_{k_i}(\Gamma)$. By considering the modular function $f_1^{k_2}/f_2^{k_1}$ of weight 0, show that

$$k_2 f_2 D(f_1) - k_1 f_1 D(f_2) \in S_{k_1+k_2+2}(\Gamma) .$$

Note that this generalizes Exercise 3.20.

2. Compute constants a, b, and c (depending on k_1 and k_2 and not all 0) such that

$$[f_1, f_2]_2 = aD^2(f_1) + bD(f_1)D(f_2) + cD^2(f_2) \in S_{k_1+k_2+4}(\Gamma) .$$

This gives the first two of the so-called *Rankin–Cohen* brackets.

As an application of derivatives of modular forms, we give a proof of a theorem of Siegel. We begin by the following:

Lemma 3.40 *Let a and b be nonnegative integers such that $4a + 6b = 12r + 2$. The constant term of the Fourier expansion of $F_r(a, b) = E_4^a E_6^b/\Delta^r$ vanishes.*

Proof By assumption $F_r(a, b)$ is a meromorphic modular form of weight 2. Since $D(\sum_{n\ge n_0} a(n)q^n) = \sum_{n\ge n_0} na(n)q^n$, it is sufficient to find a modular function $G_r(a, b)$ of weight 0 such that $F_r(a, b) = D(G_r(a, b))$ (recall that the derivative of a modular function of weight 0 is still modular). We prove this by an induction first on r, then on b. Recall that by Exercise 3.32, we have $D(j) = -E_{14}/\Delta = -E_4^2 E_6/\Delta$, and since $4a + 6b = 14$ has only the solution $(a, b) = (2, 1)$ the result is true for $r = 1$. Assume it is true for $r - 1$. We now do a recursion on b, noting that since $2a + 3b = 6r + 1$, b is odd. Note that $D(j^r) = rj^{r-1}D(j) = -rE_4^{3r-1}E_6/\Delta^r$, so the constant term of $F_r(a, 1)$ indeed vanishes. However, since $E_4^3 - E_6^2 = 1728\Delta$, if $a \ge 3$, we have

$$F_r(a-3, b+2) = E_4^{a-3}E_6^b(E_4^3 - 1728\Delta)/\Delta^r = F_r(a, b) - 1728F_{r-1}(a-3, b)\ ,$$

proving that the result is true for r by induction on b since we assumed it true for $r - 1$. □

We can now prove (part of) Siegel's theorem:

Theorem 3.41 *For* $r = \dim(M_k(\Gamma))$ *define coefficients* c_i^k *by*

$$\frac{E_{12r-k+2}}{\Delta^r} = \sum_{i \ge -r} c_i^k q^i\ ,$$

where by convention we set $E_0 = 1$. *Then, for any* $f = \sum_{n \ge 0} a(n) \in M_k(\Gamma)$, *we have the relation*

$$\sum_{0 \le n \le r} c_{-n}^k a(n) = 0\ .$$

In addition, we have $c_0^k \neq 0$, *so that* $a(0) = \sum_{1 \le n \le r}(c_{-n}^k/c_0^k)a(n)$ *is a linear combination with* rational coefficients *of the* $a(n)$ *for* $1 \le n \le r$.

Proof First note that by Corollary 3.6, we have $r \ge (k-2)/12$ (with equality only if $k \equiv 2 \pmod{12}$), so the definition of the coefficients c_i^k makes sense. Note also that since the Fourier expansion of $E_{12r-k+2}$ begins with $1 + O(q)$ and that of Δ^r by $q^r + O(q^{r+1})$, that of the quotient begins with $q^{-r} + O(q^{1-r})$ (in particular, $c_{-r}^k = 1$). The proof of the first part is now immediate: the modular form $fE_{12r-k+2}$ belongs to $M_{12r+2}(\Gamma)$, so by Corollary 3.7 is a linear combination of $E_4^a E_6^b$ with $4a + 6b = 12r + 2$. It follows from the lemma that the constant term of $fE_{12r-k+2}/\Delta^r$ vanishes, and this constant term is equal to $\sum_{0 \le n \le r} c_{-n}^k a(n)$, proving the first part of the theorem. The fact that $c_0^k \neq 0$ (which is of course essential) is a little more difficult and will be omitted, see [1] Theorem 9.5.1. □

This theorem has (at least) two consequences. First, a theoretical one: if one can construct a modular form whose constant term is some interesting quantity and whose Fourier coefficients $a(n)$ are rational, this shows that the interesting quantity is also rational. This is what allowed Siegel to show that the value at negative integers of Dedekind zeta functions of totally real number fields are rational, see Sect. 7.2. Second, a practical one: it allows to compute explicitly the constant coefficient $a(0)$ in terms of the $a(n)$, giving interesting formulas, see again Sect. 7.2.

4 Hecke Operators: Ramanujan's Discoveries

We now come to one of the most amazing and important discoveries on modular forms due to S. Ramanujan, which has led to the modern development of the subject. Recall that we set

$$\Delta(\tau) = q \prod_{m \geq 1} (1 - q^m)^{24} = \sum_{n \geq 1} \tau(n) q^n \ .$$

We have $\tau(2) = -24$, $\tau(3) = 252$, and $\tau(6) = -6048 = -24 \cdot 252$, so that $\tau(6) = \tau(2)\tau(3)$. After some more experiments, Ramanujan conjectured that if m and n are coprime, we have $\tau(mn) = \tau(m)\tau(n)$. Thus, by decomposing an integer into products of prime powers, assuming this conjecture, we are reduced to the study of $\tau(p^k)$ for p prime.

Ramanujan then noticed that $\tau(4) = -1472 = (-24)^2 - 2^{11} = \tau(2)^2 - 2^{11}$, and again after some experiments he conjectured that $\tau(p^2) = \tau(p)^2 - p^{11}$, and more generally that $\tau(p^{k+1}) = \tau(p)\tau(p^k) - p^{11}\tau(p^{k-1})$. Thus, $u_k = \tau(p^k)$ satisfies a linear recurrence relation

$$u_{k+1} - \tau(p)u_k + p^{11}u_{k-1} = 0 \ ,$$

and since $u_0 = 1$ the sequence is entirely determined by the value of $u_1 = \tau(p)$. It is well known that the behavior of a linear recurrent sequence is determined by its *characteristic polynomial*. Here, it is equal to $X^2 - \tau(p)X + p^{11}$, and the third of Ramanujan's conjectures is that the discriminant of this equation is always negative, or equivalently that $|\tau(p)| < p^{11/2}$.

Note that if α_p and β_p are the roots of the characteristic polynomial (necessarily distinct since we cannot have $|\tau(p)| = p^{11/2}$), then $\tau(p^k) = (\alpha_p^{k+1} - \beta_p^{k+1})/(\alpha_p - \beta_p)$, and the last conjecture says that α_p and β_p are *complex conjugate*, and, in particular, of modulus *equal* to $p^{11/2}$.

These conjectures are all true. The first two (multiplicativity and recursion) were proved by L. Mordell only 1 year after Ramanujan formulated them, and indeed the proof is quite easy (in fact, we will prove them below). The third conjecture $|\tau(p)| < p^{11/2}$ is extremely hard, and was only proved by P. Deligne in 1970 using the whole machinery developed by the school of A. Grothendieck to solve the Weil conjectures.

The main idea of Mordell, which was generalized later by E. Hecke, is to introduce certain linear operators (now called Hecke operators) on spaces of modular forms, to prove that they satisfy the multiplicativity and recursion properties (this is, in general, much easier than to prove this on numbers), and finally to use the fact that $S_{12}(\Gamma) = \mathbb{C}\Delta$ is of dimension 1, so that necessarily Δ is an *eigenform* of the Hecke operators whose eigenvalues are exactly its Fourier coefficients.

Although there are more natural ways of introducing them, we will define the Hecke operator $T(n)$ on $M_k(\Gamma)$ directly by its action on Fourier expansions $T(n)(\sum_{m\geq 0} a(m)q^m) = \sum_{m\geq 0} b(m)q^m$, where

$$b(m) = \sum_{d\mid\gcd(m,n)} d^{k-1}a(mn/d^2)\ .$$

Note that we can consider this definition as purely formal, apart from the presence of the integer k this is totally unrelated to the possible fact that $\sum_{m\geq 0} a(m)q^m \in M_k(\Gamma)$.

A simple but slightly tedious combinatorial argument shows that these operators satisfy

$$T(n)T(m) = \sum_{d\mid\gcd(n,m)} d^{k-1}T(nm/d^2)\ .$$

In particular, if m and n are coprime, we have $T(n)T(m) = T(nm)$ (multiplicativity), and if p is a prime and $k \geq 1$, we have $T(p^k)T(p) = T(p^{k+1}) + p^{k-1}T(p^{k-1})$ (recursion). This shows that these operators are indeed good candidates for proving the first two of Ramanujan's conjectures.

We need to show the essential fact that they preserve $M_k(\Gamma)$ and $S_k(\Gamma)$ (the latter will follow from the former since by the above definition $b(0) = \sum_{d\mid n} d^{k-1}a(0) = a(0)\sigma_{k-1}(n) = 0$ if $a(0) = 0$). By recursion and multiplicativity, it is sufficient to show this for $T(p)$ with p prime. Now, if $F(\tau) = \sum_{m\geq 0} a(m)q^m$, $T(p)(F)(\tau) = \sum_{m\geq 0} b(m)q^m$ with $b(m) = a(mp)$ if $p \nmid m$, and $b(m) = a(mp) + p^{k-1}a(m/p)$ if $p \mid m$.

On the other hand, let us compute $G(\tau) = \sum_{0\leq j<p} F((\tau + j)/p)$. Replacing directly in the Fourier expansion, we have

$$G(\tau) = \sum_{m\geq 0} a(m)q^{m/p} \sum_{0\leq j<p} e^{2\pi imj/p}\ .$$

The inner sum is a complete geometric sum which vanishes unless $p \mid m$, in which case it is equal to p. Thus, changing m into pm, we have $G(\tau) = p\sum_{m\geq 0} a(pm)q^m$. On the other hand, we have trivially $\sum_{p\mid m} a(m/p)q^m = \sum_{m\geq 0} a(m)q^{pm} = F(p\tau)$. Replacing both of these formulas in the formula for $T(p)(F)$, we see that

$$T(p)(F)(\tau) = p^{k-1}F(p\tau) + \frac{1}{p}\sum_{0\leq j<p} F\left(\frac{\tau + j}{p}\right)\ .$$

Exercise 4.1 Show more generally that

$$T(n)(F)(\tau) = \sum_{ad=n} a^{k-1}\frac{1}{d}\sum_{0\leq b<d} F\left(\frac{a\tau + b}{d}\right)\ .$$

It is now easy to show that $T(p)F$ is modular: replace τ by $\gamma(\tau)$ in the above formula and make a number of elementary manipulations to prove modularity. In fact, since Γ is generated by $\tau \mapsto \tau + 1$ and $\tau \mapsto -1/\tau$, it is immediate to check modularity for these two maps on the above formula.

As mentioned above, the proof of the first two Ramanujan conjectures is now immediate: since $T(n)$ acts on the one-dimensional space $S_{12}(\Gamma)$, we must have $T(n)(\Delta) = c \cdot \Delta$ for some constant c. Replacing in the definition of $T(n)$, we thus have for all m $c\tau(m) = \sum_{d|\gcd(n,m)} d^{11}\tau(nm/d^2)$. Choosing $m = 1$ and using $\tau(1) = 1$ shows that $c = \tau(n)$, so that

$$\tau(n)\tau(m) = \sum_{d|\gcd(n,m)} d^{11}\tau(nm/d^2)$$

which implies (and is equivalent to) the first two conjectures of Ramanujan.

Denote by $P_k(n)$ the *characteristic polynomial* of the linear map $T(n)$ on $S_k(\Gamma)$. A strong form of the so-called Maeda's conjecture states that for $n > 1$ the polynomial $P_k(n)$ is *irreducible*. This has been tested up to very large weights.

Exercise 4.2 The above proof shows that the Hecke operators also preserve the space of modular *functions*, so by Theorem 3.31, the image of $j(\tau)$ will be a rational function in j:

1. Show, for instance, that

$$T(2)(j) = j^2/2 - 744j + 81000 \quad \text{and}$$
$$T(3)(j) = j^3/3 - 744j^2 + 356652j - 12288000 \;.$$

2. Set $J = j - 744$, i.e., j with no term in q^0 in its Fourier expansion. Deduce that

$$T(2)(J) = J^2/2 - 196884 \quad \text{and}$$
$$T(3)(J) = J^3/3 - 196884J - 21493760 \;,$$

   and observe that the coefficients that we obtain are exactly the Fourier coefficients of J.
3. Prove that $T(n)(j)$ is a *polynomial* in j. Does the last observation generalize?

5 Euler Products, Functional Equations

5.1 Euler Products

The case of Δ is quite special, in that the modular form space to which it naturally belongs, $S_{12}(\Gamma)$, is only one dimensional. As can easily be seen from the dimension formula, this occurs (for cusp forms) only for $k = 12, 16, 18, 20, 22$, and 26 (there are no nonzero cusp forms in weight 14 and the space is of dimension 2 in weight 24), and thus the evident cusp forms ΔE_{k-12} for these values of k (setting $E_0 = 1$) are generators of the space $S_k(\Gamma)$, so are eigenforms of the Hecke operators and share exactly the same properties as Δ, with p^{11} replaced by p^{k-1}.

When the dimension is greater than 1, we must work slightly more. From the formulas given above, it is clear that the $T(n)$ forms a *commutative algebra* of operators on the finite-dimensional vector space $S_k(\Gamma)$. In addition, we have seen above that there is a natural *scalar product* on $S_k(\Gamma)$. One can show the not completely trivial fact that $T(n)$ is Hermitian for this scalar product, and hence, in particular, is diagonalizable. It follows by an easy and classical result of linear algebra that these operators are *simultaneously diagonalizable*, i.e., there exists a basis F_i of forms in $S_k(\Gamma)$ such that $T(n)F_i = \lambda_i(n)F_i$ for all n and i. Identifying Fourier coefficients as we have done above for Δ shows that if $F_i = \sum_{n\geq 1} a_i(n)q^n$, we have $a_i(n) = \lambda_i(n)a_i(1)$. This implies first that $a_i(1) \neq 0$, otherwise F_i would be identically zero, so that by dividing by $a_i(1)$ we can always *normalize* the eigenforms so that $a_i(1) = 1$, and second, as for Δ, that $a_i(n) = \lambda_i(n)$, i.e., the eigenvalues are exactly the Fourier coefficients. In addition, since the $T(n)$ are Hermitian, these eigenvalues are real for any embedding into $\mathbb{C}$, and hence are *totally real*, in other words their minimal polynomial has only real roots. Finally, using Theorem 3.5, it is immediate to show that the field generated by the $a_i(n)$ is finite dimensional over $\mathbb{Q}$, i.e., is a number field.

Exercise 5.1 Consider the space $S = S_{24}(\Gamma)$, which is the smallest weight where the dimension is greater than 1, here 2. By the structure theorem given above, it is generated, for instance, by Δ^2 and ΔE_4^3. Compute the matrix of the operator $T(2)$ on this basis of S, diagonalize this matrix, so find the *eigenfunctions* of $T(2)$ on S (the prime number 144169 should occur). Check that these eigenfunctions are also eigenfunctions of $T(3)$.

Thus, let $F = \sum_{n\geq 1} a(n)q^n$ be a *normalized* eigenfunction for all the Hecke operators in $S_k(\Gamma)$ (for instance, $F = \Delta$ with $k = 12$), and consider the *Dirichlet series*

$$L(F, s) = \sum_{n\geq 1} \frac{a(n)}{n^s} ,$$

for the moment formally, although we will show below that it converges for $\Re(s)$ sufficiently large. The multiplicativity property of the coefficients ($a(nm) = a(n)a(m)$ if $\gcd(n, m) = 1$, coming from that of the $T(n)$) is *equivalent* to the fact that we have an *Euler product* (a product over primes)

$$L(F, s) = \prod_{p\in P} L_p(F, s) \quad \text{with} \quad L_p(F, s) = \sum_{j\geq 0} \frac{a(p^j)}{p^{js}} ,$$

where we will always denote by P the set of prime numbers.

The additional recursion property $a(p^{j+1}) = a(p)a(p^j) - p^{k-1}a(p^{j-1})$ is equivalent to the identity

$$L_p(F,s) = \frac{1}{1 - a(p)p^{-s} + p^{k-1}p^{-2s}}$$

(multiply both sides by the denominator to check this). We have thus proved the following theorem:

Theorem 5.2 *Let $F = \sum_{n\geq 1} a(n)q^n \in S_k(\Gamma)$ be an eigenfunction of all Hecke operators. We have an Euler product*

$$L(F,s) = \sum_{n\geq 1} \frac{a(n)}{n^s} = \prod_{p\in P} \frac{1}{1 - a(p)p^{-s} + p^{k-1}p^{-2s}} .$$

Note that we have not really used the fact that F is a cusp form: the above theorem is still valid if $F = F_k$ is the normalized Eisenstein series

$$F_k(\tau) = -\frac{B_k}{2k}E_k(\tau) = -\frac{B_k}{2k} + \sum_{n\geq 1} \sigma_{k-1}(n)q^n ,$$

which is easily seen to be a normalized eigenfunction for all Hecke operators. In fact,

Exercise 5.3 Let $a \in \mathbb{C}$ be any complex number and let as usual $\sigma_a(n) = \sum_{d|n} d^a$.

1. Show that

$$\sum_{n\geq 1} \frac{\sigma_a(n)}{n^s} = \zeta(s-a)\zeta(s) = \prod_{p\in P} \frac{1}{1 - \sigma_a(p)p^{-s} + p^a p^{-2s}} ,$$

with $\sigma_a(p) = p^a + 1$.

2. Show that

$$\sigma_a(m)\sigma_a(n) = \sum_{d|\gcd(m,n)} d^a \sigma_a\left(\frac{mn}{d^2}\right) ,$$

so that, in particular, F_k is indeed a normalized eigenfunction for all Hecke operators.

5.2 *Analytic Properties of L-Functions*

Everything that we have done up to now is purely formal, i.e., we do not need to assume convergence. However, in the sequel, we will need to prove some analytic results, and for this, we need to prove convergence for certain values of s. We begin with the following easy bound, due to Hecke:

Proposition 5.4 *Let $F = \sum_{n\geq 1} a(n)q^n \in S_k(\Gamma)$ be a cusp form (not necessarily an eigenform). There exists a constant $c > 0$ (depending on F) such that for all n, we have $|a(n)| \leq cn^{k/2}$.*

Proof The trick is to consider the function $g(\tau) = |F(\tau)\Im(\tau)^{k/2}|$: since we have seen that $\Im(\gamma(\tau)) = \Im(\tau)/|c\tau + d|^2$, it follows that $g(\tau)$ is *invariant* under Γ. It follows that $\sup_{\tau\in\mathscr{H}} g(\tau) = \sup_{\tau\in\mathfrak{F}} g(\tau)$, where $\mathfrak{F}$ is the fundamental domain used above. Now because of the Fourier expansion and the fact that F is a cusp form, $|F(\tau)| = O(e^{-2\pi\Im(\tau)})$ as $\Im(\tau) \to \infty$, so $g(\tau)$ tends to 0 also. It immediately follows that g is *bounded* on $\mathfrak{F}$, and hence on $\mathscr{H}$, so that there exists a constant $c_1 > 0$ such that $|F(\tau)| \leq c_1\Im(\tau)^{-k/2}$ for all τ.

We can now easily prove Hecke's bound: from the Fourier series section, we know that for any $y > 0$

$$a(n) = e^{2\pi ny} \int_0^1 F(x + iy)e^{-2\pi inx}\, dx\ ,$$

so that $|a(n)| \leq c_1 e^{2\pi ny} y^{-k/2}$, and choosing $y = 1/n$ proves the proposition with $c = e^{2\pi} c_1$. ☐

The following corollary is now clear:

Corollary 5.5 *The L-function of a cusp form of weight k converges absolutely (and uniformly on compact subsets) for $\Re(s) > k/2 + 1$.*

Remark 5.6 Deligne's deep result mentioned above on the third Ramanujan conjecture implies that we have the following optimal bound: there exists $c > 0$ such that $|a(n)| \leq c\sigma_0(n)n^{(k-1)/2}$, and, in particular, $|a(n)| = O(n^{(k-1)/2+\varepsilon})$ for all $\varepsilon > 0$. This implies that the L-function of a cusp form converges absolutely and uniformly on compact subsets, in fact, also for $\Re(s) > (k + 1)/2$.

Exercise 5.7 Define for all $s \in \mathbb{C}$ the function $\sigma_s(n)$ by $\sigma_s(n) = \sum_{d|n} d^s$ if $n \in \mathbb{Z}_{>0}$, $\sigma_s(0) = \zeta(-s)/2$ (and $\sigma_s(n) = 0$ otherwise). Set

$$S(s_1, s_2; n) = \sum_{0\leq m\leq n} \sigma_{s_1}(m)\sigma_{s_2}(n - m)\ .$$

1. Compute $S(s_1, s_2; n)$ exactly in terms of $\sigma_{s_1+s_2+1}(n)$ for $(s_1, s_2) = (3, 3)$ and $(3, 5)$, and also for $(s_1, s_2) = (1, 1)$, $(1, 3)$, $(1, 5)$, and $(1, 7)$ by using properties of the function E_2.
2. Using Hecke's bound for cusp forms, show that if s_1 and s_2 are odd positive integers the ratio $S(s_1, s_2; n)/\sigma_{s_1+s_2+1}(n)$ tends to a limit $L(s_1, s_2)$ as $n \to \infty$, and compute this limit in terms of Bernoulli numbers. In addition, give an estimate for the *error term* $|S(s_1, s_2; n)/\sigma_{s_1+s_2+1}(n) - L(s_1, s_2)|$.
3. Using the values of the Riemann zeta function at even positive integers in terms of Bernoulli numbers, show that if s_1 and s_2 are odd positive integers, we have

$$L(s_1, s_2) = \frac{\zeta(s_1+1)\zeta(s_2+1)}{(s_1+s_2+1)\binom{s_1+s_2}{s_1}\zeta(s_1+s_2+2)} \ .$$

4. (A little project.) *Define* $L(s_1, s_2)$ by the above formula for all s_1, s_2 in $\mathbb{C}$ for which it makes sense, interpreting $\binom{s_1+s_2}{s_1}$ as $\Gamma(s_1+s_2+1)/(\Gamma(s_1+1)\Gamma(s_2+1))$. Check on a computer whether it still seems to be true that

$$S(s_1, s_2; n)/\sigma_{s_1+s_2+1}(n) \to L(s_1, s_2) \ .$$

Try to *prove* it for $s_1 = s_2 = 2$, and then for general s_1, s_2. If you succeed, give also an estimate for the error term analogous to the one obtained above.

We now do some (elementary) analysis.

Proposition 5.8 *Let $F \in S_k(\Gamma)$. For $\Re(s) > k/2+1$, we have*

$$(2\pi)^{-s}\Gamma(s)L(F, s) = \int_0^\infty F(it)t^{s-1}\,dt \ .$$

Proof Using $\Gamma(s) = \int_0^\infty e^{-t}t^{s-1}\,dt$, this is trivial by uniform convergence which insures that we can integrate term by term. □

Corollary 5.9 *The function $L(F, s)$ is a holomorphic function which can be analytically continued to the whole of $\mathbb{C}$. In addition, if we set $\Lambda(F, s) = (2\pi)^{-s}\Gamma(s)$ $L(F, s)$, we have the functional equation $\Lambda(F, k-s) = i^{-k}\Lambda(F, s)$.*

Note that in our case k is even, so that $i^{-k} = (-1)^{k/2}$, but we prefer writing the constant as above so as to be able to use a similar result in odd weight, which occurs in more general situations.

Proof Indeed, splitting the integral at 1, changing t into $1/t$ in one of the integrals, and using modularity shows immediately that

$$(2\pi)^{-s}\Gamma(s)L(F, s) = \int_1^\infty F(it)(t^{s-1} + i^k t^{k-1-s})\,dt \ .$$

Since the integral converges absolutely and uniformly for all s (recall that $F(it)$ tends exponentially fast to 0 when $t \to \infty$), this immediately implies the corollary. □

As an aside, note that the integral formula used in the above proof is a very efficient numerical method to compute $L(F, s)$, since the series obtained on the right by term-by-term integration is exponentially convergent. For instance,

Exercise 5.10 Let $F(\tau) = \sum_{n\geq 1} a(n)q^n$ be the Fourier expansion of a cusp form of weight k on Γ. Using the above formula, show that the value of $L(F, k/2)$ at the center of the "critical strip" $0 \leq \Re(s) \leq k$ is given by the following exponentially convergent series:

$$L(F,k/2) = (1+(-1)^{k/2})\sum_{n\geq 1}\frac{a(n)}{n^{k/2}}e^{-2\pi n}P_{k/2}(2\pi n)\ ,$$

where $P_{k/2}(X)$ is the polynomial

$$P_{k/2}(X) = \sum_{0\leq j<k/2} X^j/j! = 1 + X/1! + X^2/2! + \cdots + X^{k/2-1}/(k/2-1)!\ .$$

Note, in particular, that if $k \equiv 2 \pmod 4$, we have $L(F,k/2) = 0$. Prove this directly.

Exercise 5.11 1. Prove that if F is not necessarily a cusp form, we have $|a(n)| \leq cn^{k-1}$ for some $c > 0$.
2. Generalize the proposition and the integral formulas so that they are also valid for non-cusp forms; you will have to add polar parts of the type $1/s$ and $1/(s-k)$.
3. Show that $L(F,s)$ still extends to the whole of $\mathbb{C}$ with functional equation, but that it has a pole, simple, at $s=k$, and compute its residue. In passing, show that $L(F,0) = -a(0)$.

5.3 *Special Values of L-Functions*

A general "paradigm" on L-functions, essentially due to P. Deligne, is that if some "natural" L-function has both an Euler product and functional equations similar to the above, then for suitable integral "special points" the value of the L-function should be a certain (a priori transcendental) number ω times an algebraic number.

In the case of modular forms, this is a theorem of Yu. Manin.

Theorem 5.12 *Let F be a normalized eigenform in $S_k(\Gamma)$, and denote by K the number field generated by its Fourier coefficients. There exist two nonzero complex numbers ω_+ and ω_- such that for $1 \leq j \leq k-1$ integral, we have*

$$\Lambda(F,j)/\omega_{(-1)^j} \in K\ ,$$

where we recall that $\Lambda(F,s) = (2\pi)^{-s}\Gamma(s)L(F,s)$.

In addition, $\omega_\pm$ can be chosen such that $\omega_+\omega_- = < F, F >$.

In other words, for j odd, we have $L(F,j)/\omega_- \in K$ while for j even, we have $L(F,j)/\omega_+ \in K$.

For instance, in the case $F = \Delta$, if we choose $\omega_- = \Lambda(F,3)$ and $\omega_+ = \Lambda(F,2)$, we have

$$(\Lambda(F,j))_{1\leq j\leq 11\ odd} = (1620/691, 1, 9/14, 9/14, 1, 1620/691)\omega_-$$
$$(\Lambda(F,j))_{1\leq j\leq 11\ even} = (1, 25/48, 5/12, 25/48, 1)\omega_+\ ,$$

and $\omega_+\omega_- = (8192/225) < F, F >$.

Exercise 5.13 (*see also Exercise* 3.9). For $F \in S_k(\Gamma)$ define the *period polynomial* $P(F, X)$ by

$$P(F;X) = \int_0^{i\infty} (X-\tau)^{k-2} F(\tau)\, d\tau\ .$$

1. For $\gamma \in \Gamma$ show that

$$P(F;X)|_{2-k} = \int_{\gamma^{-1}(0)}^{\gamma^{-1}(i\infty)} (X-\tau)^{k-2} F(\tau)\, d\tau\ .$$

2. Show that $P(F;X)$ satisfies

$$\begin{aligned} &P(F;X)|_{2-k}S + P(F;X) = 0 \quad \text{and} \\ &P(F;X)|_{2-k}(ST)^2 + P(F;X)|_{2-k}(ST) + P(F;X) = 0\ . \end{aligned}$$

3. Show that

$$P(F;X) = -\sum_{j=0}^{k-2} (-i)^{k-1-j} \binom{k-2}{j} \Lambda(F, k-1-j) X^j\ .$$

4. If $F = \Delta$, using Manin's theorem above show that up to the multiplicative constant ω_+, $\Re(P(F;X))$ factors completely in $\mathbb{Q}[X]$ as a product of linear polynomials, and show a similar result for $\Im(P(F;X))$ after omitting the extreme terms involving 691.

5.4 Nonanalytic Eisenstein Series and Rankin–Selberg

If we replace the expression $(c\tau + d)^k$ by $|c\tau + d|^{2s}$ for some complex number s, we can also obtain functions which are invariant by Γ, although they are nonanalytic. More precisely:

Definition 5.14 Write as usual $y = \Im(\tau)$. For $\Re(s) > 1$, we define

$$G(s)(\tau) = \sum_{(c,d)\in\mathbb{Z}^2\setminus\{(0,0)\}} \frac{y^s}{|c\tau+d|^{2s}} \quad \text{and}$$

$$E(s)(\tau) = \sum_{\gamma\in\Gamma_\infty\setminus\Gamma} \Im(\gamma(\tau))^s = \frac{1}{2}\sum_{\gcd(c,d)=1} \frac{y^s}{|c\tau+d|^{2s}}\ .$$

This is again an *averaging* procedure, and it follows that $G(s)$ and $E(s)$ are *invariant* under Γ. In addition, as in the case of the holomorphic Eisenstein series

G_k and E_k, it is clear that $G(s) = 2\zeta(2s)E(s)$. One can also easily compute their Fourier expansion, and the result is as follows:

Proposition 5.15 *Set $\Lambda(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$. We have the Fourier expansion*

$$\Lambda(2s)E(s) = \Lambda(2s)y^s + \Lambda(2-2s)y^{1-s} + 4y^{1/2}\sum_{n\ge1}\frac{\sigma_{2s-1}(n)}{n^{s-1/2}}K_{s-1/2}(2\pi ny)\cos(2\pi nx)\ .$$

In the above, $K_\nu(x)$ is a K-Bessel function which we do not define here. The main properties that we need is that it tends to 0 exponentially (more precisely $K_\nu(x) \sim (\pi/(2x))^{1/2}e^{-x}$ as $x \to \infty$) and that $K_{-\nu} = K_\nu$. It follows from the above Fourier expansion that $E(s)$ has an *analytic continuation* to the whole complex plane, that it satisfies the functional equation $\mathscr{E}(1-s) = \mathscr{E}(s)$, where we set $\mathscr{E}(s) = \Lambda(2s)E(s)$, and that $E(s)$ has a unique pole, at $s = 1$, which is simple with residue $3/\pi$, independent of τ.

Exercise 5.16 Using the properties of the Riemann zeta function $\zeta(s)$, show this last property, i.e., that $E(s)$ has a unique pole, at $s = 1$, which is simple with residue $3/\pi$, independent of τ.

There are many reasons for introducing these non-holomorphic Eisenstein series, but for us the main reason is that they are fundamental in *unfolding* methods. Recall that using unfolding, in Proposition 3.12 we showed that E_k (or G_k) was orthogonal to any cusp form. In the present case, we obtain a different kind of result called a *Rankin–Selberg convolution*. Let f and g be in $M_k(\Gamma)$, one of them being a cusp form. Since $E(s)$ is invariant by Γ, the scalar product $< E(s)f, g >$ makes sense, and the following proposition gives its value:

Proposition 5.17 *Let $f(\tau) = \sum_{n\ge0} a(n)q^n$ and $g(\tau) = \sum_{n\ge0} b(n)q^n$ be in $M_k(\Gamma)$, with at least one being a cusp form. For $\Re(s) > 1$, we have*

$$< E(s)f, g >= \frac{\Gamma(s+k-1)}{(4\pi)^{s+k-1}}\sum_{n\ge1}\frac{a(n)\overline{b(n)}}{n^{s+k-1}}\ .$$

Proof We essentially copy the proof of Proposition 3.12 so we skip the details: setting temporarily $F(\tau) = f(\tau)\overline{g(\tau)}y^k$ which is invariant by Γ, we have

$$\begin{aligned}< E(s)f, g > &= \int_{\Gamma\backslash\mathscr{H}}\sum_{\gamma\in\Gamma_\infty\backslash\Gamma}\Im(\gamma(\tau))^sF(\gamma(\tau))\,d\mu\\ &= \sum_{\Gamma_\infty\backslash\mathscr{H}}\Im(\tau)^sF(\tau)\,d\mu\\ &= \int_0^\infty y^{s+k-2}\int_0^1 F(x+iy)\,dx\,dy\ .\end{aligned}$$

The inner integral is equal to the constant term in the Fourier expansion of F, and hence is equal to $\sum_{n\ge1} a(n)\overline{b(n)}e^{-4\pi ny}$ (note that by assumption one of f and g is a cusp form, so the term $n = 0$ vanishes), and the proposition follows. □

Corollary 5.18 *For $\Re(s) > k$ set*

$$R(f, g)(s) = \sum_{n\ge1} \frac{a(n)\overline{b(n)}}{n^s}\ .$$

1. *$R(f, g)(s)$ has an analytic continuation to the whole complex plane and satisfies the functional equation $\mathscr{R}(2k - 1 - s) = \mathscr{R}(s)$ with*

$$\mathscr{R}(s) = \Lambda(2s - 2k + 1)(4\pi)^{-s}\Gamma(s)R(f, g)(s)\ .$$

2. *$R(f, g)(s)$ has a single pole, which is simple, at $s = k$ with residue*

$$\frac{3}{\pi}\frac{(4\pi)^k}{(k-1)!} < f, g > \ .$$

Proof This immediately follows from the corresponding properties of $E(s)$: we have

$$\Lambda(2s - 2k + 2)(4\pi)^{-s}\Gamma(s)R(f, g)(s) = < \mathscr{E}(s - k + 1)f, g > ,$$

and the right-hand side has an analytic continuation to $\mathbb{C}$, and is invariant when changing s into $2k - 1 - s$. In addition, by the proposition $E(s - k + 1) = \mathscr{E}(s - k + 1)/\Lambda(2s - 2k + 2)$ has a single pole, which is simple, at $s = k$, with residue $3/\pi$, so $R(f, g)(s)$ also has a single pole, which is simple, at $s = k$ with residue $\dfrac{3}{\pi}\dfrac{(4\pi)^k}{(k-1)!} < f, g >$. □

It is an important fact (see Theorem 7.9 of my notes on L-functions in the present volume) that L-functions having analytic continuation and standard functional equations can be very efficiently computed at any point in the complex plane (see the note after the proof of Corollary 5.9 for the special case of $L(F, s)$). Thus, the above corollary gives a very efficient method for computing Petersson scalar products.

Note that the *holomorphic* Eisenstein series $E_k(\tau)$ can also be used to give Rankin–Selberg convolutions, but now between forms of different weights.

Exercise 5.19 Let $f = \sum_{n\ge0} a(n)q^n \in M_\ell(\Gamma)$ and $g = \sum_{n\ge0} b(n)q^n \in M_{k+\ell}(\Gamma)$, at least one being a cusp form. Using exactly the same unfolding method as in the above proposition or as in Proposition 3.12, show that

$$< E_k f, g > = \frac{(k + \ell - 2)!}{(4\pi)^{k+\ell-1}} \sum_{n\ge1} \frac{a(n)\overline{b(n)}}{n^{k+\ell-1}}\ .$$

6 Modular Forms on Subgroups of Γ

6.1 Types of Subgroups

We have used as basic definition of (weak) modularity $F|_k\gamma = F$ for all $\gamma \in \Gamma$. But there is no reason to restrict to Γ: we could very well ask the same modularity condition for some group G of transformations of $\mathscr{H}$ different from Γ.

There are many types of such groups, and they have been classified: for us, we will simply distinguish three types, with no justification. For any such group G, we can talk about a fundamental domain, similar to $\mathfrak{F}$ that we have drawn above (I do not want to give a rigorous definition here). We can distinguish essentially three types of such domains, corresponding to three types of groups.

The first type is when the domain (more precisely its closure) is *compact*: we say in that case that G is *cocompact*. It is equivalent to saying that it does not have any "cusp" such as $i\infty$ in the case of G. These groups are very important, but we will not consider them here.

The second type is when the domain is not compact (i.e., it has cusps), but it has *finite volume* for the measure $d\mu = dxdy/y^2$ on $\mathscr{H}$ defined in Exercise 3.11. Such a group is said to have finite *covolume*, and the main example is $G = \Gamma$ that we have just considered, and hence also evidently all the subgroups of Γ of *finite index*.

Exercise 6.1 Show that the covolume of the modular group Γ is finite and equal to $\pi/3$.

The third type is when the volume is infinite: a typical example is the group Γ_∞ generated by integer translations, i.e., the set of matrices $\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)$. A fundamental domain is then any vertical strip in $\mathscr{H}$ of width 1, which can trivially be shown to have infinite volume. These groups are not important (at least for us) for the following reason: they would have "too many" modular forms. For instance, in the case of Γ_∞ a "modular form" would simply be a holomorphic periodic function of period 1, and we come back to the theory of Fourier series, much less interesting.

We will, therefore, restrict to groups of the second type, which are called *Fuchsian groups of the first kind*. In fact, for this course, we will even restrict to subgroups G of Γ of *finite index*.

However, even with this restriction, it is still necessary to distinguish two types of subgroups: the so-called *congruence subgroups*, and the others, of course, called non-congruence subgroups. The theory of modular forms on non-congruence subgroups is quite a difficult subject and active research is being done on them. One annoying aspect is that they apparently do not have a theory of Hecke operators.

Thus, will restrict even more to congruence subgroups. We give the following definitions:

Definition 6.2 Let $N \geq 1$ be an integer.

1. We define

$$\Gamma(N) = \{\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma,\ \gamma \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{N}\}\ ,$$

$$\Gamma_1(N) = \{\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma,\ \gamma \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod{N}\}\ ,$$

$$\Gamma_0(N) = \{\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma,\ \gamma \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N}\}\ ,$$

where the congruences are component-wise and $*$ indicates that no congruence is imposed.

2. A subgroup of Γ is said to be a *congruence subgroup* if it contains $\Gamma(N)$ for some N, and the smallest such N is called the *level* of the subgroup.

It is clear that $\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N)$, and it is trivial to prove that $\Gamma(N)$ is normal in Γ (hence, in any subgroup of Γ containing it), that $\Gamma_1(N)/\Gamma(N) \simeq \mathbb{Z}/N\mathbb{Z}$ (with the map $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \mapsto b \bmod N$), and that $\Gamma_1(N)$ is normal in $\Gamma_0(N)$ with $\Gamma_0(N)/\Gamma_1(N) \simeq (\mathbb{Z}/N\mathbb{Z})^*$ (with the map $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \mapsto d \bmod N$).

If G is a congruence subgroup of level N, we have $\Gamma(N) \subset G$, so (whatever the definition) a modular form on G will, in particular, be on $\Gamma(N)$. Because of the above isomorphisms, it is not difficult to reduce the study of forms on $\Gamma(N)$ to those on $\Gamma_1(N)$, and the latter to forms on $\Gamma_0(N)$, except that we have to add a slight "twist" to the modularity property. Thus, for simplicity, we will restrict to modular forms on $\Gamma_0(N)$.

6.2 Modular Forms on Subgroups

In view of the definition given for Γ, it is natural to say that F is weakly modular of weight k on $\Gamma_0(N)$ if for all $\gamma \in \Gamma_0(N)$, we have $F|_k\gamma = F$, where we recall that if $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, then $F|_k\gamma(\tau) = (c\tau + d)^{-k}F(\tau)$. To obtain a modular *form*, we need also to require that F is holomorphic on $\mathcal{H}$, plus some additional technical condition "at infinity". In the case of the full modular group Γ, this condition was that $F(\tau)$ remains bounded as $\Im(\tau) \to \infty$. In the case of a subgroup, this condition is not sufficient (it is easy to show that if we do not require an additional condition the corresponding space will, in general, be infinite dimensional). There are several equivalent ways of giving the additional condition. One is the following: writing as usual $\tau = x + iy$, we require that there exists N such that in the strip $-1/2 \le x \le 1/2$, we have $|F(\tau)| \le y^N$ as $y \to \infty$ and $|F(\tau)| \le y^{-N}$ as $y \to 0$ (since F is 1-periodic, there is no loss of generality in restricting to the strip).

It is easily shown that if F is weakly modular and holomorphic, then the above inequalities imply that $|F(\tau)|$ is, in fact, *bounded* as $y \to \infty$ (but, in general, *not* as $y \to 0$), so the first condition is exactly the one that we gave in the case of the full modular group.

Similarly, we can define a *cusp form* by asking that in the above strip $|F(\tau)|$ tends to 0 as $y \to \infty$ and as $y \to 0$.

Exercise 6.3 If $F \in M_k(\Gamma)$ show that the second condition $|F(\tau)| \le y^{-N}$ as $y \to 0$ is satisfied.

Now that we have a solid definition of modular form, we can try to proceed as in the case of the full modular group. A number of things can easily be generalized. It is always convenient to choose a system of representatives (γ_j) of right cosets for $\Gamma_0(N)$ in Γ, so that

$$\Gamma = \bigsqcup_j \Gamma_0(N)\gamma_j \ .$$

For instance, if $\mathfrak{F}$ is the fundamental domain of Γ seen above, one can choose $\mathscr{D} = \bigsqcup \gamma_j(\mathfrak{F})$ as fundamental domain for $\Gamma_0(N)$. The theorem that we gave on valuations generalizes immediately:

$$\sum_{\tau \in \overline{\mathscr{D}}} \frac{v_\tau(F)}{e_\tau} = \frac{k}{12}[\Gamma : \Gamma_0(N)] \ ,$$

where $\overline{\mathscr{D}}$ is $\mathscr{D}$ to which is added a finite number of "cusps" (we do not explain this; it is *not* the topological closure), $e_\tau = 2$ (resp., 3) if τ is Γ-equivalent to i (resp., to ρ), and $e_\tau = 1$ otherwise, and we can then deduce the dimension of $M_k(\Gamma_0(N))$ and $S_k(\Gamma_0(N))$ as we did for Γ:

Theorem 6.4 *We have $M_0(\Gamma_0(N)) = \mathbb{C}$ (i.e., the only modular forms of weight 0 are the constants) and $S_0(\Gamma_0(N)) = \{0\}$. For $k \ge 2$ even, we have*

$$\begin{aligned} \dim(M_k(\Gamma_0(N))) &= A_1 - A_{2,3} - A_{2,4} + A_3 \quad \textit{and} \\ \dim(S_k(\Gamma_0(N))) &= A_1 - A_{2,3} - A_{2,4} - A_3 + \delta_{k,2} \ , \end{aligned}$$

where $\delta_{k,2}$ is the Kronecker symbol (1 if $k = 2$, 0 otherwise) and the A_i are given as follows:

$$\begin{aligned} A_1 &= \frac{k-1}{12} N \prod_{p \mid N}\left(1 + \frac{1}{p}\right) \ , \\ A_{2,3} &= \left(\frac{k-1}{3} - \left\lfloor \frac{k}{3} \right\rfloor\right) \prod_{p \mid N}\left(1 + \left(\frac{-3}{p}\right)\right) \quad \textit{if } 9 \nmid N, \quad 0 \ \textit{otherwise}, \\ A_{2,4} &= \left(\frac{k-1}{4} - \left\lfloor \frac{k}{4} \right\rfloor\right) \prod_{p \mid N}\left(1 + \left(\frac{-4}{p}\right)\right) \quad \textit{if } 4 \nmid N, \quad 0 \ \textit{otherwise}, \\ A_3 &= \frac{1}{2} \sum_{d \mid N} \phi(\gcd(d, N/d)) \ . \end{aligned}$$

6.3 Examples of Modular Forms on Subgroups

We give a few examples of modular forms on subgroups. First, note the following easy lemma:

Lemma 6.5 *If $F \in M_k(\Gamma_0(N))$, then for any $m \in \mathbb{Z}_{\geq 1}$, we have $F(m\tau) \in M_k(\Gamma_0(mN))$.*

Proof Trivial since when $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(mN)$ one can write $(m(a\tau+b)/(c\tau+d)) = (a(m\tau)+mb)/((c/m)\tau+d)$. $\square$

Thus, we can already construct many forms on subgroups, but in a sense they are not very interesting, since they are "old" in a precise sense that we will define below.

A second more interesting example is Eisenstein series: there are more general Eisenstein series than those that we have seen for Γ, but we simply give the following important example: using a similar proof to the above lemma we can construct Eisenstein series of *weight* 2 as follows. Recall that $E_2(\tau) = 1 - 24\sum_{n\geq 1}\sigma_1(n)q^n$ is not quite modular, and that $E_2^*(\tau) = E_2(\tau) - 3/(\pi\Im(\tau))$ is weakly modular (but of course non-holomorphic). Consider the function $F(\tau) = NE_2(N\tau) - E_2(\tau)$, analogous to the construction of the lemma with a correction term.

We have the evident but crucial fact that we also have $F(\tau) = NE_2^*(N\tau) - E_2^*(\tau)$ (since $\Im(\tau)$ is multiplied by N), so F is also weakly modular on $\Gamma_0(N)$, but since it is holomorphic, we have thus constructed a (nonzero) modular form of weight 2 on $\Gamma_0(N)$.

A third important example is provided by theta series. This would require a book in itself, so we restrict to the simplest case. We have seen in Corollary 1.3 that the function $T(a) = \sum_{n\in\mathbb{Z}} e^{-a\pi n^2}$ satisfies $T(1/a) = a^{1/2}T(a)$, which looks like (and is) a modularity condition. This was for $a > 0$ real. Let us generalize and for $\tau \in \mathcal{H}$ set

$$\theta(\tau) = \sum_{n\in\mathbb{Z}} q^{n^2} = \sum_{n\in\mathbb{Z}} e^{2\pi i n^2\tau} ,$$

so that, for instance, we simply have $T(a) = \theta(ia/2)$. The proof of the functional equation for T that we gave using Poisson summation is still valid in this more general case and shows that

$$\theta(-1/(4\tau)) = (2\tau/i)^{1/2}\theta(\tau) .$$

On the other hand, the definition trivially shows that $\theta(\tau+1) = \theta(\tau)$. If we denote by W_4 the matrix $\left(\begin{smallmatrix} 0 & -1 \\ 4 & 0 \end{smallmatrix}\right)$ corresponding to the map $\tau \mapsto -1/(4\tau)$ and as usual $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, we thus have $\theta|_{1/2}W_4 = c\theta$ and $\theta_{1/2}T = \theta$ for some eighth root of unity c. (Note: we always use the principal determination of the square roots; if you are uncomfortable with this, simply square everything, this is what we will do below anyway.) This implies that if we let Γ_θ be the intersection of Γ with the group generated by W_4 and T (as transformations of $\mathcal{H}$), then for all $\gamma \in \Gamma_\theta$ we will have

$\theta|_{1/2}\gamma = c(\gamma)\theta$ for some eighth root of unity $c(\gamma)$, but in fact $c(\gamma)$ is a fourth root of unity which we will give explicitly below.

One can easily describe this group Γ_θ, and in particular show that it contains $\Gamma_0(4)$ as a subgroup of index 2. This implies that $\theta^4 \in M_2(\Gamma_0(4))$, and more generally of course $\theta^{4m} \in M_{2m}(\Gamma_0(4))$.

As one of the most famous application of the finite dimensionality of modular form spaces, solve the following exercise:

Exercise 6.6 1. Using the dimension formulas, show that $2E_2(2\tau) - E_2(\tau)$ together with $4E_2(4\tau) - E_2(\tau)$ form a basis of $M_2(\Gamma_0(4))$.

2. Using the Fourier expansion of E_2, deduce an explicit formula for the Fourier expansion of θ^4, and hence that $r_4(n)$, the number of representations of n as a sum of four squares (in $\mathbb{Z}$, all permutations counted) is given for $n \geq 1$ by the formula
$$r_4(n) = 8(\sigma_1(n) - 4\sigma_1(n/4)) ,$$
where it is understood that $\sigma_1(x) = 0$ if $x \notin \mathbb{Z}$. In particular, show that this trivially implies Lagrange's theorem that every integer is a sum of four squares.

3. Similarly, show that $r_8(n)$, the nth Fourier coefficient of θ^8, is given for $n \geq 1$ by
$$r_8(n) = 16(\sigma_3(n) - 2\sigma_3(n/2) + 16\sigma_3(n/4)) .$$

Remark 6.7 Using more general methods one can give "closed" formulas for $r_k(n)$ for $k = 1, 2, 3, 4, 5, 6, 7, 8$, and 10, see, e.g., [1].

6.4 Hecke Operators and L-Functions

We can introduce the same Hecke operators as before, but to have a reasonable definition we must add a coprimality condition: we define $T(n)(\sum_{m\geq 0} a(m)q^m) = \sum_{m\geq 0} b(m)q^m$, with
$$b(m) = \sum_{\substack{d\mid\gcd(m,n)\\ \gcd(d,N)=1}} d^{k-1}a(mn/d^2) .$$

This additional condition $\gcd(d, N) = 1$ is, of course, automatically satisfied if n is coprime to N, but not otherwise.

One then shows exactly like in the case of the full modular group that
$$T(n)T(m) = \sum_{\substack{d\mid\gcd(n,m)\\ \gcd(d,N)=1}} d^{k-1}T(nm/d^2) ,$$
that they preserve modularity, so, in particular, the $T(n)$ form a commutative algebra of operators on $S_k(\Gamma_0(N))$. And this is where the difficulties specific to subgroups of

Γ begin: in the case of Γ, we stated (without proof nor definition) that the $T(n)$ were *Hermitian* with respect to the Petersson scalar product, and deduced the existence of eigenforms for *all* Hecke operators. Unfortunately, here the same proof shows that the $T(n)$ are Hermitian when n is coprime to N, but *not* otherwise.

It follows that there exist common eigenforms for the $T(n)$, but *only* for n coprime to N, which creates difficulties.

An analogous problem occurs for *Dirichlet characters*: if χ is a Dirichlet character modulo N, it may, in fact, come by natural extension from a character modulo M for some divisor $M \mid N$, $M < N$. The characters which have nice properties, in particular, with respect to the functional equation of their L-functions, are the *primitive* characters, for which such an M does not exist.

A similar but slightly more complicated thing can be done for modular forms. It is clear that if $M \mid N$ and $F \in M_k(\Gamma_0(M))$, then, of course, $F \in M_k(\Gamma_0(N))$. More generally, by Lemma 6.5, for any $d \mid N/M$, we have $F(d\tau) \in M_k(\Gamma_0(N))$. Thus, we want to exclude such "oldforms". However, it is not sufficient to say that a newform is not an oldform. The correct definition is to define a newform as a form which is *orthogonal* to the space of oldforms with respect to the scalar product, and, of course, the new space is the space of newforms. Note that in the case of Dirichlet characters, this orthogonality condition (for the standard scalar product of two characters) is automatically satisfied so need not be added.

This theory was developed by Atkin–Lehner–Li, and the new space $S_k^{\mathrm{new}}(\Gamma_0(N))$ can be shown to have all the nice properties that we require. Although not trivial, one can prove that it has a basis of common eigenforms for *all* Hecke operators, not only those with n coprime to N. More precisely, one shows that in the new space an eigenform for the $T(n)$ for all n coprime to N is automatically an eigenform for *any* operator which commutes with all the $T(n)$, such as, of course, the $T(m)$ for $\gcd(m, N) > 1$.

In addition, we have not really lost anything by restricting to the new space, since it is easy to show that

$$S_k(\Gamma_0(N)) = \bigoplus_{M \mid N} \bigoplus_{d \mid N/M} B(d) S_k^{\mathrm{new}}(\Gamma_0(M)) \ ,$$

where $B(d)$ is the operator sending $F(\tau)$ to $F(d\tau)$. Note that the sums in the above formula are *direct* sums.

Exercise 6.8 The above formula shows that

$$\dim(S_k(\Gamma_0(N))) = \sum_{M \mid N} \sigma_0(N/M) \dim(S_k^{\mathrm{new}}(\Gamma_0(M))) \ ,$$

where $\sigma_0(n)$ is the number of divisors of n.

1. Using the Möbius inversion formula, show that if we define an arithmetic function β by $\beta(p) = -2$, $\beta(p^2) = 1$, and $\beta(p^k) = 0$ for $k \geq 3$, and extend by multiplicativity ($\beta(\prod p_i^{v_i}) = \prod \beta(p_i^{v_i})$), we have the following dimension formula for the

new space:

$$\dim(S_k^{\mathrm{new}}(\Gamma_0(N))) = \sum_{M \mid N} \beta(N/M) \dim(S_k(\Gamma_0(M))) \ .$$

2. Using Theorem 6.4, deduce a direct formula for the dimension of the new space.

Proposition 6.9 *Let $F \in S_k(\Gamma_0(N))$ and $W_N = \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right)$.*

1. *We have $F|_k W_N \in S_k(\Gamma_0(N))$, where*

$$F|_k W_N(\tau) = N^{-k/2} \tau^{-k} F(-1/(N\tau)) \ .$$

2. *If F is an eigenform (in the new space), then $F|_k W_N = \pm F$ for a suitable sign $\pm$.*

Proof (1): This simply follows from the fact that W_N *normalizes* $\Gamma_0(N)$: $W_N^{-1} \Gamma_0(N) W_N = \Gamma_0(N)$ as can easily be checked, and the same result would be true for any other normalizing operator such as the *Atkin–Lehner* operators which we will not define. The operator W_N is called the *Fricke involution*.

(2): It is easy to show that W_N commutes with all Hecke operators $T(n)$ when $\gcd(n, N) = 1$, so by what we have mentioned above, if F is an eigenform in the new space, it is automatically an eigenform for W_N, and since W_N acts as an involution, its eigenvalues are ± 1. □

The eigenforms can again be normalized with $a(1) = 1$, and their L-function has an Euler product, of a slightly more general shape:

$$L(F, s) = \prod_{p \nmid N} \frac{1}{1 - a(p)p^{-s} + p^{k-1}p^{-2s}} \prod_{p \mid N} \frac{1}{1 - a(p)p^{-s}} \ .$$

Proposition 5.8 is, of course, still valid, but is not the correct normalization to obtain a functional equation. We replace it by

$$N^{s/2}(2\pi)^{-s}\Gamma(s)L(F, s) = \int_0^\infty F(it/N^{1/2})t^{s-1}\,dt \ ,$$

which, of course, is trivial from the proposition by replacing t by $t/N^{1/2}$. Indeed, thanks to the above proposition, we split the integral at $t = 1$, and using the action of W_N, we deduce the following proposition:

Proposition 6.10 *Let $F \in S_k^{new}(\Gamma_0(N))$ be an eigenform for all Hecke operators, and write $F|_k W_N = \varepsilon F$ for some $\varepsilon = \pm 1$. The L-function $L(F, s)$ extends to a holomorphic function in $\mathbb{C}$, and if we set $\Lambda(F, s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(F, s)$, we have the functional equation*

$$\Lambda(F, k - s) = \varepsilon i^{-k} \Lambda(F, s) \ .$$

Proof Indeed, the trivial change of variable t into $1/t$ proves the formula

$$N^{s/2}(2\pi)^{-s}\Gamma(s)L(F,s)=\int_1^{\infty}F(it/N^{1/2})(t^{s-1}+\varepsilon i^k t^{k-1-s})\,dt\ ,$$

from which the result follows. □

Once again, we leave to the reader to check that if $F(\tau)=\sum_{n\geq 1}a(n)q^n$, we have

$$L(F,k/2)=(1+\varepsilon(-1)^{k/2})\sum_{n\geq 1}\frac{a(n)}{n^{k/2}}e^{-2\pi n/N^{1/2}}P_{k/2}(2\pi n/N^{1/2})\ .$$

6.5 Modular Forms with Characters

Consider again the problem of sums of squares, in other words of the powers of $\theta(\tau)$. We needed to raise it to a power which is a multiple of 4 so as to have a pure modularity property as we defined it above. But consider the function $\theta^2(\tau)$. The same proof that we mentioned for θ^4 shows that for any $\gamma=\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)\in\Gamma_0(4)$, we have

$$\theta^2(\gamma(\tau))=\left(\frac{-4}{d}\right)(c\tau+d)\theta^2(\tau)\ ,$$

where $\left(\frac{-4}{d}\right)$ is the Legendre–Kronecker character (in this specific case equal to $(-1)^{(d-1)/2}$ since d is odd, being coprime to c). Thus, it satisfies a modularity property, except that it is "twisted" by $\left(\frac{-4}{d}\right)$. Note that the equation makes sense since if we change γ into $-\gamma$ (which does not change $\gamma(\tau)$), then $(c\tau+d)$ is changed into $-(c\tau+d)$, and $\left(\frac{-4}{d}\right)$ is changed into $\left(\frac{-4}{-d}\right)=-\left(\frac{-4}{d}\right)$. It is thus essential that the multiplier that we put in front of $(c\tau+d)^k$, here $\left(\frac{-4}{d}\right)$, has the same parity as k.

We mentioned above that the study of modular forms on $\Gamma_1(N)$ could be reduced to those on $\Gamma_0(N)$ "with a twist". Indeed, more precisely it is trivial to show that

$$M_k(\Gamma_1(N))=\bigoplus_{\chi(-1)=(-1)^k}M_k(\Gamma_0(N),\chi)\ ,$$

where χ ranges through all Dirichlet characters modulo N of the specified parity, and where $M_k(\Gamma_0(N),\chi)$ is defined as the space of functions F satisfying

$$F(\gamma(\tau))=\chi(d)(c\tau+d)^kF(\tau)$$

for all $\gamma=\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)\in\Gamma_0(N)$, plus the usual holomorphy and conditions at the cusps (note that $\gamma\mapsto\chi(d)$ are the group homomorphism from $\Gamma_0(N)$ to $\mathbb{C}^*$ which induces the abovementioned isomorphism from $\Gamma_0(N)/\Gamma_1(N)$ to $(\mathbb{Z}/N\mathbb{Z})^*$).

Exercise 6.11 1. Show that a system of coset representatives of $\Gamma_1(N)\backslash\Gamma_0(N)$ is given by matrices $M_d = \left(\begin{smallmatrix} u & -v \\ N & d \end{smallmatrix}\right)$, where $0 \le d < N$ such that $\gcd(d, N) = 1$ and u and v are such that $ud + vN = 1$.

2. Let $f \in M_k(\Gamma_1(N))$. Show that in the above decomposition of $M_k(\Gamma_1(N))$, we have $f = \sum_{\chi(-1)=(-1)^k} f_\chi$ with

$$f_\chi = \sum_{0\le d<N,\ \gcd(d,N)=1} \overline{\chi(d)} f|_k M_d \ .$$

These spaces are just as nice as the spaces $M_k(\Gamma_0(N))$ and share exactly the same properties. They have finite dimension (which we do not give), there are Eisenstein series, Hecke operators, newforms, Euler products, L-functions, etc. An excellent rule of thumb is simply to replace any formula containing d^{k-1} (or p^{k-1}) by $\chi(d)d^{k-1}$ (or $\chi(p)p^{k-1}$). In fact, in the Euler product of the L-function of an eigenform, we do not need to distinguish $p \nmid N$ and $p \mid N$ since we have

$$L(F, s) = \prod_{p\in P} \frac{1}{1 - a(p)p^{-s} + \chi(p)p^{k-1-2s}} \ ,$$

and $\chi(p) = 0$ if $p \mid N$ since χ is a character modulo N.

Thus, for instance, $\theta^2 \in M_1(\Gamma_0(4), \chi_{-4})$, more generally $\theta^{4m+2} \in M_{2m+1}(\Gamma_0(4), \chi_{-4})$, where we use the notation χ_D for the Legendre–Kronecker symbol $\left(\frac{D}{d}\right)$.

The space $M_1(\Gamma_0(4), \chi_{-4})$ has dimension 1, generated by the single Eisenstein series

$$1 + 4\sum_{n\ge1} \sigma_0^{(-4)}(n)q^n \ , \quad \text{where} \quad \sigma_{k-1}^{(D)}(n) = \sum_{d\mid n} \left(\frac{D}{d}\right) d^{k-1}$$

according to our rule of thumb (which does not tell us the constant 4). Comparing constant coefficients, we deduce that $r_2(n) = 4\sigma_0^{(-4)}(n)$, where as usual $r_2(n)$ is the number of representations of n as a sum of two squares. This formula was in essence discovered by Fermat.

For $r_6(n)$, we must work slightly more: $\theta^6 \in M_3(\Gamma_0(4), \chi_{-4})$, and this space has dimension 2, generated by two Eisenstein series. The first is the natural "rule of thumb" one (which again does not give us the constant)

$$F_1 = 1 - 4\sum_{n\ge1} \sigma_2^{(-4)}(n)q^n \ ,$$

and the second is

$$F_2 = \sum_{n\ge1} \sigma_2^{(-4,*)}(n)q^n \ ,$$

where

$$\sigma_{k-1}^{(D,*)}=\sum_{d\mid n}\left(\frac{D}{n/d}\right)d^{k-1}\;,$$

a sort of dual to $\sigma_{k-1}^{(D)}$ (these are my notation). Since $\theta^6 = 1 + 12q + \cdots$, comparing the Fourier coefficients of 1 and q shows that $\theta^6 = F_1 + 16F_2$, so we deduce that

$$r_6(n) = -4\sigma_2^{(-4)}(n) + 16\sigma_2^{(-4,*)}(n) = \sum_{d\mid n}\left(16\left(\frac{-4}{n/d}\right) - 4\left(\frac{-4}{d}\right)\right)d^2\;.$$

6.6 Remarks on Dimension Formulas and Galois Representations

The explicit dimension formulas alluded to above are valid for $k \in \mathbb{Z}$ *except* for $k = 1$; in addition, thanks to the theorems mentioned below, we also have explicit dimension formulas for $k \in 1/2 + \mathbb{Z}$. Thus, the theory of modular forms of weight 1 is very special, and their general construction is more difficult.

This is also reflected in the construction of *Galois representations* attached to modular eigenforms, which is an important and deep subject that we will not mention in this course, except to say the following: in weight $k \geq 2$ these representations are ℓ-adic (or modulo ℓ), i.e., with values in $\mathrm{GL}_2(\mathbb{Q}_\ell)$ (or $\mathrm{GL}_2(\mathbb{F}_\ell)$), while in weight 1 they are *complex* representations, i.e., with values in $\mathrm{GL}_2(\mathbb{C})$. The construction in weight 2 is quite old, and comes directly from the construction of the so-called *Tate module* $T(\ell)$ attached to an Abelian variety (more precisely the Jacobian of a modular curve), while the construction in higher weight, due to Deligne, is much deeper since it implies the third Ramanujan conjecture $|\tau(p)| < p^{11/2}$. Finally, the case of weight 1 is due to Deligne–Serre, in fact, using the construction for $k \geq 2$ and congruences.

6.7 Origins of Modular Forms

Modular forms are all pervasive in mathematics, physics, and combinatorics. We just want to mention the most important constructions:

- Historically, the first modular forms were probably *theta functions* (this dates back to J. Fourier at the end of the eighteenth century in his treatment of the *heat equation*) such as $\theta(\tau)$ seen above, and more generally theta functions associated with *lattices*. These functions can have integral or half-integral weight (see below) depending on whether the number of variables which occurs (equivalently, the dimension of the lattice) is even or odd. Later, these theta functions were generalized by introducing *spherical polynomials* associated with the lattice.

For example, the theta function associated to the lattice $\mathbb{Z}^2$ is simply $f(\tau) = \sum_{(x,y)\in\mathbb{Z}^2} q^{x^2+y^2}$, which is clearly equal to θ^2, so belongs to $M_1(\Gamma_0(4), \chi_{-4})$. But, we can also consider, for instance

$$f_5(\tau) = \sum_{(x,y)\in\mathbb{Z}^2} (x^4 - 6x^2y^2 + y^4)q^{x^2+y^2} ,$$

and show that $f_5 \in S_5(\Gamma_0(4), \chi_{-4})$.

Exercise 6.12 1. Using the notation and results of Exercise 3.39, show that $[\theta, \theta]_2 = cf_5$ for a suitable constant c, so that, in particular, $f_5 \in S_5(\Gamma_0(4), \chi_{-4})$.
2. Show that the polynomial $P(x, y) = x^4 - 6x^2y^2 + y^4$ is a *spherical polynomial*, in other words that $D(P) = 0$, where D is the Laplace differential operator $D = \partial^2/\partial^2 x + \partial^2/\partial^2 y$.

- The second occurrence of modular forms is probably *Eisenstein series*, which, in fact, are the first that we encountered in this course. We have only seen the most basic Eisenstein series G_k (or normalized versions) on the full modular group and a few on $\Gamma_0(4)$, but there are very general constructions over any space such as $M_k(\Gamma_0(N), \chi)$. Their Fourier expansions can easily be explicitly computed and are similar to what we have given above. More difficult is the case when k is only half-integral, but this can also be done.
  As we have seen, an important generalization of Eisenstein series is *Poincaré series*, which can also be defined over any space as above.
- A third important construction of modular forms comes from the Dedekind eta function $\eta(\tau)$ defined above. In itself it has a complicated *multiplier system*, but if we define an *eta quotient* as $F(\tau) = \prod_{m\in I} \eta(m\tau)^{r_m}$ for a certain set I of positive integers and exponents $r_m \in \mathbb{Z}$, then it is not difficult to write necessary and sufficient conditions for F to belong to some $M_k(\Gamma_0(N), \chi)$. The first example that we have met is, of course, the Ramanujan delta function $\Delta(\tau) = \eta(\tau)^{24}$. Other examples are, for instance, $\eta(\tau)\eta(23\tau) \in S_1(\Gamma_0(23), \chi_{-23})$, $\eta(\tau)^2\eta(11\tau)^2 \in S_2(\Gamma_0(11))$, and $\eta(2\tau)^{30}/\eta(\tau)^{12} \in S_9(\Gamma_0(8), \chi_{-4})$.
- Closely related to eta-quotients are q-identities involving the q-Pochhammer symbol $(q)_n$ and generalizing those seen in Exercise 3.30, many of which give modular forms not related to the eta function.
- A much deeper construction comes from algebraic geometry: by the modularity theorem of Wiles et al., to any elliptic curve defined over $\mathbb{Q}$ is associated a modular form in $S_2(\Gamma_0(N))$ which is a normalized Hecke eigenform, where N is the so-called *conductor* of the curve. For instance, the eta quotient of level 11 just seen above is the modular form associated to the isogeny class of the elliptic curve of conductor 11 with equation $y^2 + y = x^3 - x^2 - 10x - 20$.

7 More General Modular Forms

In this brief section, we will describe modular forms of a more general kind than those seen up to now.

7.1 Modular Forms of Half-Integral Weight

Coming back again to the function θ, the formulas seen above suggest that θ itself must be considered a modular form, of weight $1/2$. We have already mentioned that

$$\theta^2(\gamma(\tau)) = \left(\frac{-4}{d}\right)(c\tau+d)\theta^2(\tau)\ .$$

But what about θ itself? For this, we must be very careful about the determination of the square root:

Notation: $z^{1/2}$ will *always* denote the principal determination of the square root, i.e., such that $-\pi/2 < \mathrm{Arg}(z^{1/2}) \le \pi/2$. For instance, $(2i)^{1/2} = 1+i$, $(-1)^{1/2} = i$. Warning: we do not, in general, have $(z_1z_2)^{1/2} = z_1^{1/2}z_2^{1/2}$, but only up to sign. As a second notation, when k is odd, $z^{k/2}$ will *always* denote $(z^{1/2})^k$ and *not* $(z^k)^{1/2}$ (for instance, $(2i)^{3/2} = (1+i)^3 = -2+2i$, while $((2i)^3)^{1/2} = 2-2i$).

Thus, let us try and take the square root of the modularity equation for θ^2:

$$\theta(\gamma(\tau)) = v(\gamma,\tau)\left(\frac{-4}{d}\right)^{1/2}(c\tau+d)^{1/2}\ ,$$

where $v(\gamma,\tau) = \pm 1$ and may depend on γ and τ. A detailed study of Gauss sums shows that $v(\gamma,\tau) = \left(\frac{-4c}{d}\right)$, the general Kronecker symbol, so that the modularity equation for θ is, for any $\gamma \in \Gamma_0(4)$:

$$\theta(\gamma(\tau)) = v_\theta(\gamma)(c\tau+d)^{1/2}\theta(\tau) \quad \text{with} \quad v_\theta(\gamma) = \left(\frac{c}{d}\right)\left(\frac{-4}{d}\right)^{-1/2}\ .$$

Note that there is something very subtle going on here: this complicated *theta multiplier system* $v_\theta(\gamma)$ must satisfy a complicated *cocycle relation* coming from the trivial identity $\theta((\gamma_1\gamma_2)(\tau)) = \theta(\gamma_1(\gamma_2(\tau)))$ which can be shown to be equivalent to the general *quadratic reciprocity law*.

The following definition is due to G. Shimura:

Definition 7.1 Let $k \in 1/2+\mathbb{Z}$. A function F from $\mathscr{H}$ to $\mathbb{C}$ will be said to be a modular form of (half-integral) weight k on $\Gamma_0(N)$ with character χ if for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$, we have

$$F(\gamma(\tau)) = v_\theta(\gamma)^{2k}\chi(d)(c\tau+d)^k F(\tau)\ ,$$

and if the usual holomorphy and conditions at the cusps are satisfied (equivalently if $F^2 \in M_{2k}(\Gamma_0(N), \chi^2\chi_{-4})$).

Note that if $k \in 1/2 + \mathbb{Z}$, we have $v_\theta(\gamma)^{4k} = \chi_{-4}$, which explains the extra factor χ_{-4} in the above definition.

Since $v_\theta(\gamma)$ is defined only for $\gamma \in \Gamma_0(4)$, we need $\Gamma_0(N) \subset \Gamma_0(4)$, in other words $4 \mid N$. In addition, by definition $v_\theta(\gamma)(c\tau + d)^{1/2} = \theta(\gamma(\tau))/\theta(\tau)$ is invariant if we change γ into $-\gamma$, so if $k \in 1/2 + \mathbb{Z}$, the same is true of $v_\theta(\gamma)^{2k}(c\tau + d)^k$, and hence it follows that in the above definition, we must have $\chi(-d) = \chi(d)$, i.e., χ must be an *even* character ($\chi(-1) = 1$).

As usual, we denote by $M_k(\Gamma_0(N), \chi)$ and $S_k(\Gamma_0(N), \chi)$, the spaces of modular and cusp forms. The theory is more difficult than the theory in integral weight, but is now well developed. We mention a few items as follows:

1. There is an explicit but more complicated *dimension formula* due to J. Oesterlé and the author.
2. By a theorem of Serre–Stark, modular forms of weight $1/2$ are simply linear combinations of *unary theta functions* generalizing the function θ above.
3. One can easily construct Eisenstein series, but the computation of their Fourier expansion, due to Shimura and the author, is more complicated.
4. As usual, if we can express θ^m solely in terms of Eisenstein series, this leads to explicit formulas for $r_m(n)$, the number of representation of n as a sum of m squares. Thus, we obtain explicit formulas for $r_3(n)$ (due to Gauss), $r_5(n)$ (due to Smith and Minkowski), and $r_7(n)$, so if we complement the formulas in integral weight, we have explicit formulas for $r_m(n)$ for $1 \le m \le 8$ and $m = 10$.
5. The deeper part of the theory, which is specific to the half-integral weight case, is the existence of *Shimura lifts* from $M_k(\Gamma_0(N), \chi)$ to $M_{2k-1}(\Gamma_0(N/2), \chi^2)$, the description of the *Kohnen subspace* $S_k^+(\Gamma_0(N), \chi)$ which allows both the Shimura lift to go down to level $N/4$, and also to define a suitable Atkin–Lehner type new space, and the deep results of Waldspurger, which nicely complement the work of Shimura on lifts.

We could try to find other types of interesting modularity properties than those coming from θ. For instance, we have seen that the Dedekind eta function is a modular form of weight $1/2$ (not in Shimura's sense), and more precisely it satisfies the following modularity equation, now for any $\gamma \in \Gamma$:

$$\eta(\gamma(\tau)) = v_\eta(\gamma)(c\tau + d)^{1/2}\eta(\tau) \ ,$$

where $v_\eta(\gamma)$ is a very complicated 24-th root of unity. We could, of course, define η-modular forms of half-integral weight $k \in 1/2 + \mathbb{Z}$ by requiring $F(\gamma(\tau)) = v_\eta(\gamma)^{2k}(c\tau + d)^k F(\tau)$, but it can be shown that this would not lead to any interesting theory (more precisely the only interesting functions would be *eta-quotients* $F(\tau) = \prod_m \eta(m\tau)^{r_m}$, which can be studied directly without any new theory.

Note that there are functional relations between η and θ.

Proposition 7.2 *We have*

$$\theta(\tau) = \frac{\eta^2(\tau+1/2)}{\eta(2\tau+1)} = \frac{\eta^5(2\tau)}{\eta^2(\tau)\eta^2(4\tau)}\ .$$

Exercise 7.3 1. Prove these relations in the following way: first show that the right-hand sides satisfy the same modularity equations as θ for $T = \left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and $W_4 = \left(\begin{smallmatrix}0&-1\\4&0\end{smallmatrix}\right)$, so, in particular, that they are weakly modular on $\Gamma_0(4)$, and second show that they are really modular forms, in other words, that they are holomorphic on $\mathscr{H}$ and at the cusps.
2. Using the definition of η, deduce two *product expansions* for $\theta(\tau)$.

We could also try to study modular forms of fractional or even real weight k not integral or half-integral, but this would lead to functions with no interesting *arithmetical* properties.

In a different direction, we can relax the condition of holomorphy (or meromorphy) and ask that the functions be eigenfunctions of the *hyperbolic Laplace operator*

$$\Delta = -y^2\left(\frac{\partial^2}{\partial^2 x} + \frac{\partial^2}{\partial^2 y}\right) = -4y^2\frac{\partial^2}{\partial\tau\partial\overline{\tau}}$$

which can be shown to be invariant under Γ (more generally, under $\mathrm{SL}_2(\mathbb{R})$) together with suitable boundedness conditions. This leads to the important theory of *Maass forms*. The case of the eigenvalue 0 reduces to ordinary modular forms since $\Delta(F) = 0$ is equivalent to F being a linear combination of a holomorphic and antiholomorphic (i.e., conjugate to a holomorphic) function, each of which will be modular or conjugate of modular.

The case of the eigenvalue $1/4$ also leads to functions having nice arithmetical properties, but all other eigenvalues give functions with (conjecturally) transcendental coefficients, but these functions are useful in number theory for other reasons which we cannot explain here. Note that a famous conjecture of Selberg asserts that for *congruence subgroups*, there are no eigenvalues λ with $0 < \lambda < 1/4$. For instance, for the full modular group, the smallest nonzero eigenvalue is $\lambda = 91.1412\cdots$, which is quite large.

Exercise 7.4 Using the fact that Δ is invariant under Γ, show that $\Delta(\Im(\gamma(\tau))) = s(1-s)\Im(\gamma(\tau))$ and deduce that the non-holomorphic Eisenstein series $E(s)$ introduced in Definition 5.14 is an eigenfunction of the hyperbolic Laplace operator with eigenvalue $s(1-s)$ (note that it does not satisfy the necessary boundedness conditions, so it is not a Maass form: the functions $E(s)$ with $\Re(s) = 1/2$ constitute what is called the *continuous spectrum*, and the Maass forms the *discrete spectrum* of Δ acting on $\Gamma\backslash\mathscr{H}$).

7.2 Modular Forms in Several Variables

The last generalization that we want to mention (there are much more!) is to several variables. The natural idea is to consider holomorphic functions from $\mathscr{H}^r$ to $\mathbb{C}$, now for some $r > 1$, satisfying suitable modularity properties. If we simply ask that $\gamma \in \Gamma$ (or some subgroup) acts component-wise, we will not obtain anything interesting. The right way to do it, introduced by Hilbert–Blumenthal, is to consider a *totally real* number field K of degree r, and denote by Γ_K the group of matrices $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}_{\mathrm{K}})$, where $\mathbb{Z}_K$ is the ring of algebraic integers of K (we could also consider the larger group $\mathrm{GL}_2(\mathbb{Z}_{\mathrm{K}})$, which leads to a very similar theory). Such a γ has r *embeddings* γ_i into $\mathrm{SL}_2(\mathbb{R})$, which we will denote by $\gamma_i = \left(\begin{smallmatrix} a_i & b_i \\ c_i & d_i \end{smallmatrix}\right)$, and the correct definition is to ask that

$$F(\gamma_1(\tau_1), \cdots, \gamma_r(\tau_r)) = (c_1\tau_1 + d_1)^k \cdots (c_r\tau_r + d_r)^k F(\tau_1, \ldots, \tau_r) \ .$$

Note that the restriction to totally real number fields is due to the fact that for γ_i to preserve the upper half-plane, it is necessary that $\gamma_i \in \mathrm{SL}_2(\mathbb{R})$. Note also that the γ_i are *not* independent, they are conjugates of a single $\gamma \in \mathrm{SL}_2(\mathbb{Z}_{\mathrm{K}})$.

A holomorphic function satisfying the above is called a *Hilbert–Blumenthal* modular form (of *parallel weight k*, one can also consider forms where the exponents for the different embeddings are not equal), or more simply a Hilbert modular form (note that there are no "conditions at infinity", since one can prove that they are automatically satisfied unless $K = \mathbb{Q}$).

Since $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}_{\mathrm{K}})$ is equal to all its conjugates, such modular forms have Fourier expansions, but using the action of $\left(\begin{smallmatrix} 1 & \alpha \\ 0 & 1 \end{smallmatrix}\right)$ with $\alpha \in \mathbb{Z}_K$, it is easy to show that these expansions are of a special type, involving the *codifferent* $\mathfrak{d}^{-1}$ of K, which is the fractional ideal of $x \in K$ such that $\mathrm{Tr}(x\mathbb{Z}_K) \subset \mathbb{Z}$, where Tr denotes the trace.

One can construct Eisenstein series, here called Hecke–Eisenstein series, and compute their Fourier expansion. One of the important consequences of this computation is that it gives an explicit formula for the value $\zeta_K(1-k)$ of the *Dedekind zeta function* of K at negative integers (hence, by the functional equation of ζ_K, also at positive even integers), and, in particular, it proves that these values are *rational numbers*, a theorem due to C.-L. Siegel as an immediate consequence of Theorem 3.41. An example is as follows:

Proposition 7.5 *Let $K = \mathbb{Q}(\sqrt{D})$ be a real quadratic field with D a fundamental discriminant. Then*

1. *We have*

$$\zeta_K(-1) = \frac{1}{60} \sum_{|s|<\sqrt{D}} \sigma_1\left(\frac{D-s^2}{4}\right) ,$$

$$\zeta_K(-3) = \frac{1}{120} \sum_{|s|<\sqrt{D}} \sigma_3\left(\frac{D-s^2}{4}\right) .$$

2. *We also have formulas such as*

$$\sum_{|s|<\sqrt{D}} \sigma_1(D-s^2) = 60\left(9-2\left(\frac{D}{2}\right)\right)\zeta_K(-1)\;,$$

$$\sum_{|s|<\sqrt{D}} \sigma_3(D-s^2) = 120\left(129-8\left(\frac{D}{2}\right)\right)\zeta_K(-3)\;.$$

We can, of course, reformulate these results in terms of L-functions by using $L(\chi_D,-1) = -12\zeta_K(-1)$ and $L(\chi_D,-3) = 120\zeta_K(-3)$, where as usual χ_D is the quadratic character modulo D.

Exercise 7.6 Using Exercise 6.6 and the above formulas, show that the number $r_5(D)$ of representations of D as a sum of five squares is given by

$$r_5(D) = 480\left(5-2\left(\frac{D}{2}\right)\right)\zeta_K(-1) = -40\left(5-2\left(\frac{D}{2}\right)\right)L(\chi_D,-1)\;.$$

Note that this formula can be generalized to arbitrary D, and is due to Smith and (much later) to Minkowski. There also exists a similar formula for $r_7(D)$: when $-D$ (*not* D) is a fundamental discriminant

$$r_7(D) = -28\left(41-4\left(\frac{D}{2}\right)\right)L(\chi_{-D},-2)\;.$$

Note also that if we restrict to the *diagonal* $\tau_1 = \cdots = \tau_r$, a Hilbert modular form of (parallel) weight k gives rise to an ordinary modular form of weight kr.

We finish this section with some terminology with no explanation: if K is *not* a totally real number field, one can also define modular forms, but they will not be defined on products of the upper half-plane $\mathscr{H}$ alone, but will also involve the *hyperbolic* 3-*space* $\mathscr{H}_3$. Such forms are called *Bianchi* modular forms.

A different generalization, close to the Weierstrass $\wp$-function seen above, is the theory of *Jacobi forms*, due to M. Eichler and D. Zagier. One of the many interesting aspects of this theory is that it mixes in nontrivial way properties of forms of integral weight with forms of half-integral weight.

Finally, we mention *Siegel modular forms*, introduced by C.-L. Siegel, which are defined on higher dimensional *symmetric spaces*, on which the *symplectic groups* $\mathrm{Sp}_{2n}(\mathbb{R})$ act. The case $n = 1$ gives ordinary modular forms, and the next simplest, $n = 2$, is closely related to Jacobi forms since the Fourier coefficients of Siegel modular forms of degree 2 can be expressed in terms of Jacobi forms.

8 Some Pari/GP Commands

There exist three software packages which are able to compute with modular forms: `magma`, `Sage`, and `Pari/GP` since the spring of 2018. We give here some basic `Pari/GP` commands with little or no explanation (which is available by typing `?` or `??`): we encourage the reader to read the tutorial `tutorial-mf` available with the distribution and to practice with the package since it is an excellent way to learn about modular forms. All commands begin with the prefix `mf`, with the exception of `lfunmf` which more properly belongs to the L-function package.

Creation of modular forms: `mfDelta` (Ramanujan Delta), `mfTheta` (ordinary theta function), `mfEk` (normalized Eisenstein series E_k), more generally `mfeisenstein`, `mffrometaquo` (eta-quotients), `mffromqf` (theta function of lattices with or without spherical polynomial), `mffromell` (from elliptic curves over $\mathbb{Q}$), etc...

Arithmetic operations: `mfcoefs` (Fourier coefficients at infinity), `mflinear` (linear combination, so including addition/subtraction and scalar multiplication), `mfmul`, `mfdiv`, `mfpow` (clear), etc...

Modular operations: `mfbd`, `mftwist`, `mfhecke`, `mfatkin`, `mfderivE2`, `mfbracket`, etc...

Creation of modular form *spaces*: `mfinit`, `mfdim` (dimension of the space), `mfbasis` (random basis of the space), `mftobasis` (decomposition of a form on the `mfbasis`), `mfeigenbasis` (basis of normalized eigenforms).

Searching for modular forms with given Fourier coefficients:
`mfeigensearch`, `mfsearch`.

Expansion of $F|_k\gamma$: `mfslashexpansion`.

Numerical functions: `mfeval` (evaluation at a point in $\mathscr{H}$ or at a cusp), `mfcuspval` (valuation at a cusp), `mfsymboleval` (computation of integrals over paths in the completed upper half-plane), `mfpetersson` (Petersson scalar product), `lfunmf` (L-function associated to a modular form), etc...

Note that for now `Pari/GP` is the only package for which these last functions (beginning with `mfslashexpansion`) are implemented.

9 Suggestions for Further Reading

The literature on modular forms is vast, so I will only mention the books which I am familiar with and that in my opinion will be very useful to the reader. Note that the classic book [4] is absolutely remarkable, but may be difficult for a beginning course.

In addition to the recent book [1] by Strömberg and the author (which, of course, I strongly recommend!!!), I also highly recommend the paper [5], which is essentially a small book. Perhaps the most classical reference is [3]. The more recent book [2] is more advanced since its ultimate goal is to explain the modularity theorem of Wiles et al.

References

1. H. Cohen and F. Strömberg, *Modular Forms: A Classical Approach*, Graduate Studies in Math. **179**, American Math. Soc., (2017).
2. F. Diamond and J. Shurman, *A first course in modular forms*, Graduate Texts in Math. **228**, Springer (2005),
3. T. Miyake, *Modular Forms*, Springer (1989).
4. G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Publ. Math. Soc. Japan **11**, Princeton University Press (1994) (reprinted from the 1971 original).
5. D. Zagier, *Elliptic modular forms and their applications*, in "The 1-2-3 of modular forms", Universitext, Springer (2008), pp. 1–103.