Modular Forms of One Variable

Don Zagier

Notes based on a course given in Utrecht, Spring 1991

Table of Contents

Chapter 0. The Notion of Modular Forms and a Survey of the Main Examples

The word "modular" refers to the moduli space of complex curves (= Riemann surfaces) of genus 1. Such a curve can be represented as $\mathbb{C}/\Lambda$ where $\Lambda \subset \mathbb{C}$ is a lattice, two lattices Λ_1 and Λ_2 giving rise to the same curve if $\Lambda_2 = \lambda\Lambda_1$ for some non-zero complex number λ. A **modular function** assigns to each lattice Λ a complex number $F(\Lambda)$ with $F(\Lambda_1) = F(\Lambda_2)$ if $\Lambda_2 = \lambda\Lambda_1$. Since any lattice $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ is equivalent to a lattice of the form $\mathbb{Z}\tau + \mathbb{Z}$ with τ $(= \omega_1/\omega_2)$ a non-real complex number, the function F is completely specified by the values $f(\tau) = F(\mathbb{Z}\tau + \mathbb{Z})$ with τ in $\mathbb{C}\setminus\mathbb{R}$ or even, since $f(\tau) = f(-\tau)$, with τ in the complex upper half-plane $\mathfrak{H} = \{\, \tau \in \mathbb{C} \mid \Im(\tau) > 0\}$. The fact that the lattice Λ is not changed by replacing the basis $\{\omega_1, \omega_2\}$ by the new basis $a\omega_1 + b\omega_2, c\omega_1 + d\omega_2$ $(a, b, c, d \in \mathbb{Z},\ ad - bc = \pm 1)$ translates into the **modular invariance property** $f(\frac{a\tau + b}{c\tau + d}) = f(\tau)$. Requiring that τ always belong to $\mathfrak{H}$ is equivalent to looking only at bases $\{\omega_1, \omega_2\}$ which are oriented (i.e. $\Im(\omega_1/\omega_2) > 0$) and forces us to look only at matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $ad - bc = +1$; the group $SL_2(\mathbb{Z})$ of such matrices will be denoted Γ_1 and called the (full) **modular group**. Thus a modular function can be thought of as a complex-valued function on $\mathfrak{H}$ which is invariant under the action $\tau \mapsto (a\tau + b)/(c\tau + d)$ of Γ_1 on $\mathfrak{H}$. Usually we are interested only in functions which are also holomorphic on $\mathfrak{H}$ (and satisfy a suitable growth condition at infinity) and will reserve the term "modular function" for these. The prototypical example is the **modular invariant** $j(\tau) = e^{-2\pi i\tau} + 744 + 196884e^{2\pi i\tau} + \cdots$ (cf. 1.2).

However, it turns out that for many purposes the condition of modular invariance is too restrictive. Instead, one must consider functions on lattices which satisfy the identity $F(\Lambda_1) = \lambda^k F(\Lambda_2)$ when $\Lambda_2 = \lambda\Lambda_1$ for some integer k, called the **weight**. Again the function F is completely determined by its restriction $f(\tau)$ to lattices of the form $\mathbb{Z}\tau + \mathbb{Z}$ with τ in $\mathfrak{H}$, but now f must satisfy the **modular transformation property**

$$f\left(\frac{a\tau + b}{c\tau + d}\right) = (c\tau + d)^k f(\tau) \tag{1}$$

rather than the modular invariance property required before. The advantage of allowing this more general transformation property is that now there are functions satisfying it which are not only holomorphic in $\mathfrak{H}$, but also "holomorphic at infinity" in the sense that their absolute value is majorized by a polynomial in $\max\{1, \Im(\tau)^{-1}\}$. This cannot happen for non-constant Γ_1-invariant functions by Liouville's theorem (the function $j(\tau)$ above, for instance, grows exponentially as $\Im(\tau)$ tends to infinity). Holomorphic functions $f : \mathfrak{H} \to \mathbb{C}$ satisfying (1) and the growth condition just given are called **modular forms** of weight k, and the set of all such functions—clearly a vector space over $\mathbb{C}$—is denoted by M_k or $M_k(\Gamma_1)$. The subspace of functions whose absolute value is majorized by a multiple of $\Im(\tau)^{-k/2}$ is denoted by S_k or $S_k(\Gamma_1)$, the space of **cusp forms** of weight k. One also looks at the spaces $M_k(\Gamma)$ and $S_k(\Gamma)$, defined as above but with (1) only required to hold for

$\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$, for subgroups $\Gamma \subset \Gamma_1$ of finite index, e.g. the subgroup $\Gamma_0(N)$ consisting of all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1$ with c divisible by a fixed integer N.

The definition of modular forms which we have just given may not at first look very natural. The importance of modular forms stems from the conjunction of the following two facts:

(i) They arise naturally in a wide variety of contexts in mathematics and physics and often encode the arithmetically interesting information about a problem.

(ii) The space M_k is finite-dimensional for each k.

The point is that if $\dim M_k = d$ and we have more than d situations giving rise to modular forms in M_k, then we automatically have a linear relation among these functions and hence get "for free" information—often highly non-trivial—relating these different situations. The way the information is "encoded" in the modular forms is via the Fourier coefficients. From the property (1) applied to the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ we find that any modular form $f(\tau)$ is invariant under $\tau \mapsto \tau + 1$ and hence, since it is also holomorphic, has a Fourier expansion as $\sum a_n e^{2\pi i n \tau}$. The growth conditions defining M_k and S_k as given above are equivalent to the requirement that a_n vanish for $n < 0$ or $n \le 0$, respectively (this is the form in which these growth conditions are usually stated). What we meant by (i) above is that nature—both physical and mathematical—often produces situations described by numbers which turn out to be the Fourier coefficients of a modular form. These can be as disparate as multiplicities of energy levels, numbers of vectors in a lattice of given length, sums over the divisors of integers, special values of zeta functions, or numbers of solutions of Diophantine equations. But the fact that all of these different objects land in the little spaces M_k forces the existence of relations among their coefficients. To give the flavour of the kind of results one can obtain, we will now list some of the known constructions of modular forms and some of the number-theoretic identities one obtains by studying the relations among them. More details and many more examples will be given in the following chapters.

Eisenstein series. For each integer $k \ge 2$, we have a function $G_k(\tau)$ with the Fourier development

$$G_k(\tau) = c_k + \sum_{n=1}^{\infty} \sigma_{k-1}(n)\, e^{2\pi i n \tau} ,$$

where $\sigma_{k-1}(n)$ for $n > 0$ denotes the sum of the $(k-1)$st powers of the (positive) divisors of n and c_k is a certain rational number, e.g.,

$$\begin{aligned}
G_2(\tau) &= -\frac{1}{24} + q + 3q^2 + 4q^3 + 7q^4 + 6q^5 + 12q^6 + 8q^7 + 15q^8 + \cdots \\
G_4(\tau) &= \frac{1}{240} + q + 9q^2 + 28q^3 + 73q^4 + 126q^5 + 252q^6 + \cdots \\
G_6(\tau) &= -\frac{1}{504} + q + 33q^2 + 244q^3 + 1057q^4 + \cdots \\
G_8(\tau) &= \frac{1}{480} + q + 129q^2 + 2188q^3 + \cdots
\end{aligned}$$

(here and from now on we use q to denote $e^{2\pi i\tau}$). We will study these functions in Chapter 1 and show that G_k for $k > 2$ is a modular form of weight k, while G_2 is "nearly" a modular form of weight 2 (for instance, $G_2(\tau) - NG_2(N\tau)$ is a modular form of weight 2 on $\Gamma_0(N)$ for all N). This will immediately have arithmetic consequences of interest. For instance, because the space of modular forms of weight 8 is one-dimensional, the forms $120\,G_4(\tau)^2$ and $G_8(\tau)$, which both have weight 8 and constant term $1/480$, must coincide, leading to the far from obvious identity

$$\sigma_7(n) = \sigma_3(n) + 120\sum_{m=1}^{n-1}\sigma_3(m)\sigma_3(n-m) \qquad (n > 0), \tag{2}$$

the first example of the above-mentioned phenomenon that the mere existence of modular forms, coupled with the finite-dimensionality of the spaces in which they lie, gives instant proofs of non-trivial number-theoretical relations. Another number-theoretical application of the G_k is connected with the constant terms c_k, which turn out to be essentially equal to the values of the Riemann zeta-function $\zeta(s) = \sum_{n=1}^{\infty}\frac{1}{n^s}$ at $s = k$, so that one obtains a new proof of and new insight into the identities

$$\zeta(2) = \frac{\pi^2}{6}, \quad \zeta(4) = \frac{\pi^4}{90}, \quad \zeta(6) = \frac{\pi^6}{945} \quad \dots$$

proved by Euler.

The discriminant function. This is the function defined by the infinite product

$$\Delta(\tau) = q\prod_{n=1}^{\infty}(1-q^n)^{24} = q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 - 6048q^6 + \cdots.$$

We will show in Chapter 1 that it is a cusp form of weight 12. This at once leads to identities like $1728\Delta = (240G_4)^3 - (504G_6)^2$ which let us express the coefficients of Δ in terms of the elementary number-theoretic functions $\sigma_{k-1}(n)$. More important, it provides the first example of one of the most interesting and important properties of modular forms. Namely, the coefficients of Δ are *multiplicative*, e.g., the coefficient -6048 of q^6 in the above expansion is the product of the coefficients -24 and 252 of q^2 and q^3 (more generally, the coefficient of q^{mn} is the product of the coefficients of q^m and q^n whenever m and n are coprime). This property, which was observed by Ramanujan in 1916 and proved by Mordell the following year, was developed by Hecke into the theory of Hecke operators, which is at the center of the whole theory of modular forms and will be discussed in detail in Chapter 2.

Theta series. Consider the *Jacobi theta function*

$$\theta(\tau) = \sum_{m\in\mathbb{Z}} q^{m^2} = 1 + 2q + 2q^4 + 2q^9 + \cdots.$$

The powers of $\theta(\tau)$ tell us the number of ways of representing an integer as a sum of a given number of squares, e.g.,

$$\theta(\tau)^4 = \sum_{m_1,\,m_2,\,m_3,\,m_4} q^{m_1^2+m_2^2+m_3^2+m_4^2} = 1 + \sum_{n=1}^{\infty} r_4(n)\, q^n$$

where $r_4(1) = 8$, $r_4(2) = 24$, $r_4(3) = 24$, $\ldots$ denote the number of ways of representing 1, 2, 3, $\ldots$ as $m_1^2 + m_2^2 + m_3^2 + m_4^2$ with $m_i \in \mathbb{Z}$. We will study $\theta(\tau)$ and similar functions in Chapter 3 and see that $\theta(\tau)^4$ is a modular form on $\Gamma_0(4)$ of weight 2. The theory of modular forms tells us that the vector space $M_2(\Gamma_0(4))$ is two-dimensional, spanned by $G_2(\tau) - 2G_2(2\tau)$ and $G_2(\tau) - 4G_2(4\tau)$. Hence $\theta(\tau)^4$ is a linear combination of these two elements; comparing the first two coefficients, we find $\theta(\tau)^4 = 8\big(G_2(\tau) - 4G_4(4\tau)\big)$ or

$$r_4(n) = 8 \sum_{\substack{d\mid n \\ d\not\equiv 0 \pmod 4}} d \qquad (n > 0)\,,$$

a famous formula due to Jacobi. In particular, $r_4(n) \geq 8$ for all n, so that we get an immediate proof of Lagrange's theorem that every positive integer is a sum of four squares. Similarly, the space $M_4(\Gamma_0(4))$ containing $\theta(\tau)^8$ is 3-dimensional with basis $G_4(\tau)$, $G_4(2\tau)$ and $G_4(4\tau)$ and we get the formula

$$r_8(n) = 16 \sum_{\substack{d\mid n \\ d\not\equiv 2 \pmod 4}} d^3 + 12 \sum_{\substack{d\mid n \\ d\equiv 2 \pmod 4}} d^3$$

for the number of representations of a positive integer n as a sum of 8 squares.

More generally, instead of the forms $m_1^2 + \ldots m_{2k}^2$ we can consider *any* positive definite quadratic form $Q(m_1, \ldots, m_{2k}) = \sum_{i\leq j} a_{ij} m_i m_j$ ($a_{ij} \in \mathbb{Z}$) in an even number of variables. Then the *theta series*

$$\Theta_Q(\tau) = \sum_{m_1,\,\ldots\,,\,m_{2k}} q^{Q(m_1,\ldots,m_{2k})} = 1 + \sum_{n=1}^{\infty} r_Q(n)\, q^n\ ,$$

where $(m_1, \ldots, m_{2k})$ in the first sum runs over all $(2k)$-tuples of integers and $r_Q(n)$ in the second denotes the number of integral representations of a positive integer n by the form Q, turns out to be a modular form of weight k on some group $\Gamma_0(N)$ and we can use the theory of modular forms to get information on the representation numbers $r_Q(n)$. This is the most powerful tool known for studying quadratic forms and has applications in the theory of higher-dimensional lattices, coding theory, etc.

Yet more generally, for certain polynomials $P(m_1, \ldots, m_{2k})$ (spherical homogeneous polynomials; see Chapter 3), the sum

$$\Theta_{Q,P}(\tau) = \sum_{m_1,\,\ldots\,,\,m_{2k}} P(m_1, \ldots, m_{2k})\, q^{Q(m_1,\ldots,m_{2k})} = 1 + \sum_{n=1}^{\infty} r_{Q,P}(n)\, q^n\ ,$$

turns out to be a cusp form of weight $k+d$, where d is the degree of P, so that we get information about the "weighted" representation numbers $r_{Q,P}(n)$. For example, if $Q(m_1,m_2)=m_1^2+m_2^2$ and $P(m_1,m_2)=m_1^4-6m_1^2m_2^2+m_2^4$ ($k=1$, $d=4$), then $\Theta_{Q,P}=4\bigl(q-4q^2+16q^4-14q^5+\dots\bigr)$ belongs to a space of cusp forms of weight 5 which one can show is one-dimensional and spanned by the form $\bigl(\Delta(\tau)^2\Delta(2\tau)\Delta(4\tau)^2\bigr)^{1/12}$, so we get the identity

$$\sum_{a,b\in\mathbb{Z}}\frac{a^4-6a^2b^2+b^4}{4}\,q^{a^2+b^2}=q\prod_{n=1}^{\infty}\bigl(1-q^n\bigr)^{\mathrm{g.c.d.}(n,4)+2}\,. \tag{3}$$

As an example of a spherical theta series of a more general kind (on the full modular group $\Gamma=\Gamma_1$) we mention Freeman Dyson's identity

$$\Delta(\tau)=\sum_{\substack{(x_1,\dots,x_5)\in\mathbb{Z}^5\\ x_1+\dots+x_5=0\\ x_i\equiv i\pmod 5}}\left(\frac{1}{288}\prod_{1\le i<j\le 5}(x_i-x_j)\right)q^{(x_1^2+x_2^2+x_3^2+x_4^2+x_5^2)/10}$$

for the discriminant function $\Delta(\tau)$.

Eisenstein series of half-integral weight. In considering theta series, there was no reason to look only at quadratic forms in an even number of variables. If we take the simplest possible quadratic form $Q(m_1)=m_1^2$, then the associated theta-series is just Jacobi's theta function $\theta(\tau)$, the fourth root of our first example, and as such satisfies the transformation equation

$$\theta\left(\frac{a\tau+b}{c\tau+d}\right)=\epsilon\,(c\tau+d)^{\frac12}\theta(\tau)\quad\forall\binom{a\ b}{c\ d}\in\Gamma_0(4)$$

for a certain number $\epsilon=\epsilon_{c,d}$ satisfying $\epsilon^4=1$ (ϵ can be given explicitly in terms of the Kronecker symbol $\bigl(\frac{c}{d}\bigr)$). We say that θ is a **modular form of weight $\frac12$**. More generally, we can define and study modular forms of any half-integral weight $r+\frac12$ ($r\in\mathbb{N}$). This will be done in Chapter 7. Just as for forms of integral weight, we can define Eisenstein series, although both the definition and the calculation of the Fourier coefficients become more complicated. It turns out that there exists for each $r>1$ an Eisenstein series $G_{r+\frac12}(\tau)$ of weight $r+\frac12$ whose Fourier coefficients are equal (up to elementary factors) to the values at $s=r$ of the Dedekind zeta-functions $\zeta_K(s)$ of real (for r even) or imaginary (for r odd) quadratic fields. As in our previous examples, we can get a simple formula for any $G_{r+\frac12}$ and hence in one blow a formula for the values of $\zeta_K(r)$, numbers which are of great number-theoretical interest, for all real or imaginary quadratic fields K. For instance, the coefficient of q^D in

$$G_{\frac92}(\tau)=\frac{1}{240}+\frac{1}{120}q+\frac{121}{120}q^4+2q^5+11q^8+\frac{2161}{120}q^9+46q^{12}+58q^{13}+\cdots$$

is $\dfrac{135D^{7/2}}{2\pi^8}\,\zeta_K(4)$ if D $(= 5,\ 8,\ 12,\ \dots)$ is the discriminant of a real quadratic field K; on the other hand, on comparing $G_{9/2}$ with the modular form $G_4(4\tau)\theta(\tau)$ of the same weight we find that they are equal and hence that we have the closed formula

$$\zeta_K(4) = \frac{2\pi^8}{135\,D^3\sqrt{D}} \sum_{\substack{|m|<\sqrt{D} \\ m^2\equiv D \pmod 4}} \sigma_3\Big(\frac{D-m^2}{4}\Big) \qquad (D = \text{discriminant of } K)$$

for every real quadratic field K, a fairly deep result which it is much harder to prove directly. For $r = 1$ the function $G_{3/2}$ has coefficients which are the class numbers of imaginary quadratic fields. Like G_2, it is no longer quite modular but has a "nearly modular" property which can be used to just as good effect. In this way also class numbers can be brought into the theory of modular forms, most notably in connection with the Eichler-Selberg formula for traces of Hecke operators (Chapter 8).

New forms from old. There are various methods which can be used to manufacture new modular forms out of previously constructed ones. The first and most obvious method is *multiplication*: the product of a modular form of weight k and one of weight l is a modular form of weight $k+l$. Of course we have already used this above when we compared G_4^2 and G_8 or when we expressed Δ as a linear combination of G_4^3 and G_6^2. The multiplicative property of modular forms means that the set $M_* = \bigoplus M_k$ of all modular forms on Γ_1 forms a graded ring. We will see in Chapter 1 that it in fact coincides with the ring of all polynomials in G_4 and G_6.

More interesting structures, which we will also study in Chapter 1, are connected with the possibility of obtaining new modular forms by combining derivatives of modular forms of lower weight. Given two modular forms f and g of weight k and l, respectively, there is for each positive integer ν a combination of the products of derivatives $f^{(i)}(\tau)g^{(\nu-i)}(\tau))$ $(0 \le i \le \nu)$ which is a cusp form of weight $k+l+2\nu$. For instance ($k = 4$, $l = 6$, $\nu = 1$), the function $2G_4(\tau)G_6'(\tau) - 3G_4'(\tau)G_6(\tau)$ is a cusp form of weight 12 and hence, since the space of such cusp forms is 1-dimensional, a multiple of Δ. More generally, we will see that the "extended" ring of modular forms generated by G_4, G_6, and the "near"-modular form G_2 is closed under differentiation, with the consequence that any modular form satisfies a non-linear third-order differential equation.

Finally, we can get new modular forms from old ones by looking at combinations of the functions $(c\tau+d)^{-k}f\big(\frac{a\tau+b}{c\tau+d}\big)$ ("slash operator"), where f is a modular form of weight k and $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ a 2×2 matrix with rational coefficients. Examples are the operators $f(\tau) \mapsto f(a\tau)$ (like the functions $G_4(2\tau)$ and $G_4(4\tau)$ used above in connection with $\theta(\tau)^8$) and the Hecke operators which were mentioned in connection with the function $\Delta(\tau)$.

Modular forms coming from algebraic geometry and number theory. Certain power series $\sum a(n)q^n$ whose coefficients $a(n)$ are defined by counting the number of points of algebraic varieties over finite fields are known or conjectured to be modular forms. For example, the famous "Taniyama-Weil conjecture" says that to any elliptic curve defined over $\mathbb{Q}$ there is associated a modular form $\sum a(n)q^n$ of weight 2 on some group $\Gamma_0(N)$ such that $p+1-a(p)$ for every prime number p is equal to the number of points of the elliptic curve over $\mathbb{F}_p$ (e.g., to the number of solutions of $x^3-ky^3\equiv 1$ modulo p, where k is a fixed integer). Of course, this cannot really be considered as a way of constructing modular forms, since one can usually only *prove* the modularity of the function in question if one has an independent, analytic construction of it. The Taniyama-Weil conjecture is very deep and in particular is known to imply Fermat's last theorem!

In a similar vein, one can get modular forms from algebraic number theory by looking at Fourier expansions $\sum a(n)q^n$ whose associated Dirichlet series $\sum a(n)n^{-s}$ are zeta functions coming from number fields or their characters. For instance, a theorem of Deligne and Serre says that one can get all modular forms of weight 1 in this way from the Artin L-series of two-dimensional Galois representations with odd determinant satisfying Artin's conjecture (that the L-series is holomorphic). Again, however, the usual way of applying such a result is to construct the modular form independently and then deduce that the corresponding Artin L-series satisfies Artin's conjecture.

In one situation, the situation of so-called "CM" (complex multiplication) forms, the analytic, algebraic geometric, and number theoretic approaches come together: analytically, these are the theta series $\Theta_{Q,P}$ associated to a *binary* quadratic form Q and an arbitrary spherical function P on $\mathbb{Z}^2$; geometrically, they arise from elliptic curves having complex multiplication (i.e., non-trivial endomorphisms); and number theoretically, they are given by Fourier developments whose associated Dirichlet series are the L-series of algebraic Hecke grossencharacters over an imaginary quadratic field. An example is the function (3) above. The characteristic property of CM forms is that they have highly lacunary Fourier developments, because binary quadratic forms represent only a thin subset of all integers (at most $\mathrm{O}\big(x/(\log x)^{1/2}\big)$ integers $\le x$).

Modular forms in several variables. Finally, modular forms in one variable can be obtained by restricting in some way different kinds of modular forms in more than one variable (Jacobi, Hilbert, Siegel, ...), these in turn being constructed by the appropriate generalization of one of the methods described above. Such forms will be discussed in Part II.

Exercises

1. Check that the expression $\Im(\tau)^{k/2}|f(\tau)|$ for a function $f:\mathfrak{H}\to\mathbb{C}$ satisfying (1) is invariant under $\tau\mapsto(a\tau+b)/(c\tau+d)$.

2. Show that the group Γ_1 is generated by the elements $T=\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and $S=\left(\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right)$. (Hint: Show that $a^2+b^2+c^2+d^2$ for a matrix $\gamma=\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\ne\pm\mathrm{I}_2$ in Γ_1 can be reduced by multiplying γ on the right by ST^n for a suitable integer n.) Deduce that a function $f:\mathfrak{H}\to\mathbb{C}$ belongs to $M_k(\Gamma_1)$ if and only if it has a convergent Fourier expansion $f(\tau)=\sum_{n=0}^{\infty}a(n)\,e^{2\pi in\tau}$ and satisfies the functional equation $f(-1/\tau)=\tau^k f(\tau)$.

3. Show that if $f(\tau)$ is a modular form of weight k on Γ_1 then $f(r\tau)$ is a modular form of weight k on $\Gamma_0(N)$ for any positive integers r and N with N divisible by r.

4*. Give an elementary proof of (2).

In this chapter we study the space of modular forms on the full modular group $\Gamma_1 = SL_2(\mathbb{Z})$. In particular, we prove the modularity properties of the Eisenstein series G_k and discriminant function Δ mentioned in Chapter 0, describe the structure of the ring of modular forms, and discuss the modularity properties of derivatives of modular forms. Many of the ideas (e.g., the construction of Eisenstein series) go through almost unchanged for other groups Γ, but restricting attention to Γ_1 we can simplify many of the details and give more complete results.

1.1. Eisenstein series. Recall that a modular form of weight k on Γ_1 is a function $f : \mathfrak{H} \to \mathbb{C}$ having a Fourier expansion $f(\tau) = \sum_{n=0}^{\infty} a(n)\, q^n$ $(q = e^{2\pi i\tau})$ which converges for all $\tau \in \mathfrak{H}$ and satisfying the modular transformation property

$$f\bigl(\gamma\tau\bigr) = (c\tau+d)^k\, f(\tau) \qquad \Bigl(\forall\gamma = \bigl(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\bigr) \in \Gamma_1,\ \gamma\tau = \frac{a\tau+b}{c\tau+d}\Bigr) \tag{1}$$

By Exercise 2 of Chapter 0, it is enough to require (1) for the matrix $\gamma = \bigl(\begin{smallmatrix} 0 & -1 \\ 1 & 0\end{smallmatrix}\bigr)$ only.

The first construction of such a form is a very simple one, but already here the Fourier coefficients will turn out to give interesting arithmetic functions. For k even and greater than 2, define the **Eisenstein series** of weight k by

$$G_k(\tau) = \frac{(k-1)!}{2(2\pi i)^k} \sum_{m,n}{}' \frac{1}{(m\tau+n)^k}, \tag{2}$$

where the sum is over all pairs of integers (m, n) except $(0, 0)$. (The reason for the normalizing factor $(k-1)!/2(2\pi i)^k$, which is not always included in the definition, will become clear in a moment.) This transforms like a modular form of weight k because replacing $G_k(\tau)$ by $(c\tau+d)^{-k} G_k(\frac{a\tau+b}{c\tau+d})$ simply replaces (m, n) by $(am+cn, bm+dn)$ and hence permutes the terms of the sum. We need the condition $k > 2$ to guarantee the absolute convergence of the sum (and hence the validity of the argument just given) and the condition k even because the series with k odd are identically zero (the terms (m, n) and $(-m, -n)$ cancel).

To see that G_k satisfies the growth condition defining M_k, and to have our first example of an arithmetically interesting modular form, we must compute the Fourier development. We begin with the **Lipschitz formula**

$$\sum_{n\in\mathbb{Z}} \frac{1}{(z+n)^k} \;=\; \frac{(-2\pi i)^k}{(k-1)!} \sum_{r=1}^{\infty} r^{k-1} e^{2\pi i r z} \qquad (k \in \mathbb{Z}_{\geq 2},\ z \in \mathfrak{H}),$$

which is proved in Appendix A1. Splitting the sum defining G_k into the terms with $m = 0$ and the terms with $m \neq 0$, and using the evenness of k to restrict to the terms with n positive in the first and m positive in the second case, we find

$$\begin{aligned} G_k(\tau) &= \frac{(k-1)!}{(2\pi i)^k} \sum_{n=1}^{\infty} \frac{1}{n^k} + \sum_{m=1}^{\infty} \biggl(\frac{(k-1)!}{(2\pi i)^k} \sum_{n\in\mathbb{Z}} \frac{1}{(m\tau+n)^k} \biggr) \\ &= \frac{(-1)^{k/2}(k-1)!}{(2\pi)^k} \zeta(k) + \sum_{m=1}^{\infty} \sum_{r=1}^{\infty} r^{k-1} e^{2\pi i r m\tau}, \end{aligned}$$

where $\zeta(s) = \sum\limits_{n=1}^{\infty} \dfrac{1}{n^s}$ is Riemann's zeta function. The number $\frac{(-1)^{k/2}(k-1)!}{(2\pi)^k}\zeta(k)$ is rational and in fact equals $-\dfrac{B_k}{2k}$, where B_k denotes the kth Bernoulli number (= coefficient of $\dfrac{x^k}{k!}$ in the Taylor expansion of $\dfrac{x}{e^x-1}$ around $x=0$); it is also equal to $\frac{1}{2}\zeta(1-k)$, where the definition of $\zeta(s)$ is extended to negative s by analytic continuation (cf. Appendix A2). Putting this into the formula for G_k and collecting for each n the terms with $rm = n$, we find finally

$$(3) \qquad G_k(\tau) = -\frac{B_k}{2k} + \sum_{n=1}^{\infty} \sigma_{k-1}(n)\, q^n = \frac{1}{2}\zeta(1-k) + \sum_{n=1}^{\infty} \sigma_{k-1}(n)\, q^n,$$

where $\sigma_{k-1}(n)$ as in Chapter 0 denotes $\sum_{r|n} r^{k-1}$ (sum over all positive divisors r of n) and we have again used the abbreviation $q = e^{2\pi i\tau}$. The beginnings of the Fourier developments of G_4, G_6 and G_8 were given in Chapter 0; the Fourier expansions of G_{10}, G_{12}, G_{14} and G_{16} begin

$$G_{10}(\tau) = -\frac{1}{264} + q + 513q^2 + \cdots, \qquad G_{12}(\tau) = \frac{691}{65520} + q + 2049q^2 + \cdots,$$
$$G_{14}(\tau) = -\frac{1}{24} + q + 8193q^2 + \cdots, \qquad G_{16}(\tau) = \frac{3617}{8160} + q + 32767q^2 + \cdots$$

The fact that the Fourier coefficients occurring are all rational numbers is a special case of the phenomenon that M_k in general is spanned by forms with rational Fourier coefficients. It is this phenomenon which is responsible for the richness of the arithmetic applications of the theory of modular forms.

The right-hand side of (3) makes sense also for $k = 2$ (B_2 is equal to $\frac{1}{6}$) and will be used to define a function $G_2(\tau)$. It is not a modular form (indeed, there can be no non-zero modular form f of weight 2 on the full modular group, as we will see in the next section). However, its transformation properties under the modular group can be easily determined using **Hecke's trick**: Define a function G_2^* by

$$G_2^*(\tau) = \frac{-1}{8\pi^2} \lim_{\epsilon \searrow 0} \Bigg({\sum_{m,n}}' \frac{1}{(m\tau+n)^2 |m\tau+n|^\epsilon} \Bigg).$$

The absolute convergence of the expression in parentheses for $\epsilon > 0$ shows that G_2^* transforms according to (1) (with $k = 2$), while applying the Poisson summation formula to this expression first and then taking the limit $\epsilon \searrow 0$ leads to the Fourier development

$$(4) \qquad G_2^*(\tau) = G_2(\tau) + \frac{1}{8\pi v} \qquad (\tau = u + iv \in \mathfrak{H}).$$

The fact that the non-holomorphic function G_2^* transforms like a modular form of weight 2 then implies that the holomorphic function G_2 transforms according to

$$(5) \qquad G_2\Big(\frac{a\tau+b}{c\tau+d}\Big) = (c\tau+d)^2 G_2(\tau) - \frac{c(c\tau+d)}{4\pi i} \qquad \Big(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_1\Big).$$

1.2. The discriminant function. We define a function Δ in $\mathfrak{H}$ by

$$\Delta(\tau) = q \prod_{r=1}^{\infty} (1-q^r)^{24} \qquad (\tau \in \mathfrak{H},\ q = e^{2\pi i \tau}). \tag{6}$$

Then

$$\begin{aligned}
\frac{\Delta'(\tau)}{\Delta(\tau)} &= \frac{d}{d\tau}\Big(2\pi i \tau + 24 \sum_{r=1}^{\infty} \log(1-q^r)\Big) \\
&= 2\pi i \Big(1 \ - \ 24 \sum_{r=1}^{\infty} \frac{r q^r}{1-q^r}\Big) \\
&= -48\pi i \Big(-\frac{1}{24} + \sum_{n=1}^{\infty} \big(\sum_{r|n} r\big) q^n\Big) \ = \ -48\pi i G_2(\tau).
\end{aligned}$$

The transformation formula (5) gives

$$\frac{1}{(c\tau+d)^2} \frac{\Delta'(\frac{a\tau+b}{c\tau+d})}{\Delta(\frac{a\tau+b}{c\tau+d})} \ = \ \frac{\Delta'(\tau)}{\Delta(\tau)} + 12 \frac{c}{c\tau+d}$$

or

$$\frac{d}{d\tau}\big(\log \Delta(\frac{a\tau+b}{c\tau+d})\big) = \frac{d}{d\tau} \log\big(\Delta(\tau)(c\tau+d)^{12}\big).$$

Integrating, we deduce that $\Delta(\frac{a\tau+b}{c\tau+d})$ equals a constant times $(c\tau+d)^{12}\Delta(\tau)$. Moreover, this constant must always be 1 since it is 1 for the special matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ (compare Fourier developments!) and $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ (take $\tau = i$!) and these matrices generate Γ_1. Thus $\Delta(\tau)$ satisfies equation (1) with $k = 12$. Multiplying out the product in (6) gives the expansion

$$\Delta(\tau) = q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 - 6048q^6 + 8405q^7 - \cdots \tag{7}$$

in which only positive exponents of q occur. Hence Δ is a cusp form of weight 12. The coefficient of q^n in the expansion (7) is usually denoted $\tau(n)$ and called the **Ramanujan function**; as already mentioned in Chapter 0, it is a multiplicative function of n.

Using Δ, we can determine the space of modular forms of all weights. Indeed, there can be no non-constant modular form of weight 0 (it would be a non-constant holomorphic function on the compact Riemann surface $\mathfrak{H}/\Gamma_1 \cup \{\infty\}$), and it follows that there can be no non-zero modular form of negative weight (if f had weight $m < 0$, then $f^{12}\Delta^{|m|}$ would have weight 0 and a Fourier expansion with no constant term). Also, $M_k = \{0\}$ for k odd (take $a = d = -1,\ b = c = 0$ in (1)). Furthermore, the space M_2 is also trivial, since if $f(\tau)$ were an element of M_2 then $f(\tau)\,d\tau$ would be a meromorphic differential form on the Riemann surface $\mathfrak{H}/\Gamma_1 \cup \{\infty\}$ of genus 0 with a single pole of order ≤ 1, contradicting the residue theorem. (For a less fancy version of this argument, see Exercise 2.) For k even and greater than 2, we have the direct sum decomposition $M_k = \langle G_k \rangle \oplus S_k$, since the Eisenstein series G_k has non-vanishing constant term and therefore subtracting a suitable multiple of it from an arbitrary modular form of weight k produces a form with zero constant term. Finally, S_k is isomorphic to M_{k-12}: given any cusp form f of weight k, the quotient f/Δ transforms like a modular form of weight $k-12$, is holomorphic in $\mathfrak{H}$ (since the product expansion (6) shows that Δ does not vanish there), and has a Fourier expansion with only nonnegative powers of q (since f has an expansion starting with a strictly positive power of q and Δ an expansion starting with q^1). It follows that M_k has finite dimension given by

k	<0	0	2	4	6	8	10	12	14	16	18	...	k	...	$k+12$	...
$\dim M_k$	0	1	0	1	1	1	1	2	1	2	2	...	d	...	$d+1$	...

It also follows, since both G_k and Δ have rational coefficients, that M_k has a basis consisting of forms with rational coefficients, as claimed previously; such a basis is for instance the set of monomials $\Delta^l G_{k-12l}$ with $0 \le l \le (k-4)/12$, together with the function $\Delta^{k/12}$ if k is divisible by 12. We also get the first examples of the phenomenon described in the introductory chapter that non-trivial arithmetic identities can be obtained "for free" from the finite-dimensionality of M_k. Thus both G_4^2 and G_8 belong to the one-dimensional space M_8, so they must be proportional and we obtain the identity (2) of Chapter 0. Similarly, the one-dimensionality of M_{10} and M_{14} imply the proportionality of G_4G_6 and G_{10} and of G_4G_{10}, G_6G_8, and G_{14} and hence various other identities of the same type. In fact, one deduces easily from what has just been said that *every modular form on $SL_2(\mathbb{Z})$ is uniquely expressible as a polynomial in G_4 and G_6*, i.e., the graded ring $M_* = \bigotimes M_k$ coincides with the polynomial algebra $C[G_4, G_6]$.

Comparing the Fourier expansions of the first few G_k as given in Chapter 0 and Section 1.1 and the dimensions of the first few M_k as given above, we notice that S_k is empty exactly for those values of k for which the constant term $-B_k/2k$ of G_k is the reciprocal of an integer (namely, for $k = 2$, 4, 6, 8, 10 and 14). This is not a coincidence: one knows for reasons going well beyond the scope of these lectures that, if there are cusp forms of weight k, there must always be congruences between some cusp form and the Eisenstein series of this weight. If this congruence is modulo a prime p, then p must divide the numerator of the constant term of G_k (since the constant term of the cusp form congruent to G_k modulo p is zero). Conversely, for any prime p dividing the numerator of the constant term of G_k, there is a congruence between G_k and some cusp form. As an example, for $k = 12$ the numerator of the constant term of G_k is the prime number 691 and we have the congruence $G_{12} \equiv \Delta \pmod{691}$ (e.g. $2049 \equiv -24 \pmod{691}$) due to Ramanujan.

Finally, the existence of Δ allows us to define the function

$$\begin{aligned} j(\tau) = \frac{(240G_4)^3}{\Delta} &= \frac{(1 + 240q + 2160q^2 + 6720q^3 + \cdots)^3}{q - 24q^2 + 252q^3 + 1472q^4 - \cdots} \\ &= q^{-1} + 744 + 196884q + 21493760q^2 + \cdots \end{aligned}$$

and see (since G_4^3 and Δ are modular forms of the same weight on Γ_1) that it is invariant under the action of Γ_1 on $\mathfrak{H}$. Conversely, if $\phi(\tau)$ is any modular function on $\mathfrak{H}$ which grows at most exponentially as $\Im(\tau) \to \infty$, then the function $f(\tau) = \phi(\tau)\Delta(\tau)^m$ transforms like a modular form of weight $12m$ and (if m is large enough) is bounded at infinity, so that $f \in M_{12m}$; by what we saw above, f is then a homogeneous polynomial of degree m in G_4^3 and Δ, so $\phi = f/\Delta^m$ is a polynomial of degree $\le m$ in j. This justifies calling $j(\tau)$ "the" modular invariant function. In fact, we can say more. Define a subset $\mathcal{F}$ of $\mathfrak{H}$ as the set of $\tau = u + iv$ satisfying $|u| \le \frac{1}{2}$, $|\tau| \ge 1$. Using Exercise 2 of Chapter 0, one can show that every point of $\mathfrak{H}$ can be mapped by an element of Γ_1 to a point of $\mathcal{F}$. If f is a modular form of weight k which does not vanish at infinity (i.e., has constant term $a(0) \ne 0$) or anywhere on the boundary $\partial\mathcal{F}$ of $\mathcal{F}$, then by integrating $f'(\tau)/f(\tau)$ around $\partial\mathcal{F}$ we find that f has exactly $k/12$ zeroes in the interior of $\mathcal{F}$ (in particular, k must be divisible by 12; modular forms on Γ_1 must have zeroes on the boundary of $\mathcal{F}$ and more specifically at the point i if $4 \not| k$ and at the points $(\pm 1 + i\sqrt{3})/2$ if $3 \not| k$, as one sees by applying a variant of the same argument or directly as in Exercise 3). In particular, for each $\lambda \in \mathbb{C}$ the function $(240G_4)^3 - \lambda\Delta(\tau)$ has exactly one zero in the interior of $\mathcal{F}$ if it does not vanish on $\partial\mathcal{F}$. This shows that $j(\tau)$ gives an isomorphism from the Riemann surface X obtained by identifying the edges of $\mathcal{F}$ via $\tau \sim \tau + 1$ for $u = -\frac{1}{2}$ and $\tau \sim -1/\tau$ for $|\tau| = 1$ to the set of complex numbers. Alternatively, one can observe that the function $j - \lambda$ on the closed surface obtained by adding a "point at infinity" to X has exactly one pole, at infinity, and hence exactly one zero, at some finite point. It follows that any meromorphic function on $\mathfrak{H}$ which is Γ_1-invariant and grows at most exponentially in $v = \Im(\tau)$ as $v \to \infty$ is a rational function of $j(\tau)$.

1.3. Modular forms and differential operators. The process of differentiation disturbs the property of modularity. If we start with a function f satisfying (1), then we find by differentiation that the derivative f' satisfies the equation

$$f'\Big(\frac{a\tau + b}{c\tau + d}\Big) = (c\tau + d)^{k+2}\, f'(\tau) + kc(c\tau + d)^{k+1}\, f(\tau)\,, \tag{8}$$

i.e., it behaves nearly like a modular form of weight $k + 2$ but with a slight perturbation. There are two ways to get around this difficulty and produce modular forms of higher weight by a differentiation property. The first is to combine derivatives of two different modular forms. For instance, if $g \in M_l$ is a second modular form, then writing down the analogue of (8) for g we see that the perturbations of the modularity property of $f'(\tau)g(\tau)$ and of $f(\tau)g'(\tau)$ are the same up to a factor k/l and hence that the function

$$F_1(f, g)(\tau) = \frac{1}{2\pi i}\big(l\, f'(\tau)\, g(\tau) - k\, f(\tau)\, g'(\tau)\big) \tag{9}$$

is a modular form (in fact, a cusp form, since the constant term is obviously zero) of weight $k + l + 2$. As the simplest numerical example, the function $F_1(G_4, G_6) = -\frac{1}{35}q - \cdots$ is a cusp form of weight 12 and hence equals $-\frac{1}{35}\Delta$, which gives the formula

$$\tau(n) = \frac{5\sigma_3(n) + 7\sigma_5(n)}{12}\, n \; - 35 \sum_{\substack{a,b>0 \\ a+b=n}} (6a - 4b)\sigma_3(a)\sigma_5(b)$$

for the coefficient $\tau(n)$ of q^n in Δ. The operation F_1 is antisymmetric in its two arguments and satisfies the Jacobi identity, so that it makes $M_{*-2} = \bigoplus_n M_{n-2}$ into a graded Lie algebra. More generally, for each $\nu > 0$ we have **H. Cohen's differential operator** F_ν which is defined for $f \in M_k$, $g \in M_l$ by

$$(10) \qquad F_\nu(f,g)(\tau) = (2\pi i)^{-\nu} \sum_{\mu=0}^{\nu} (-1)^\mu \binom{k+\nu-1}{\mu} \binom{l+\nu-1}{\nu-\mu} f^{(\nu-\mu)}(\tau)\, g^{(\mu)}(\tau)$$

(the normalization has been chosen so that $F_\nu(f,g)$ has integral Fourier coefficients if both f and g do). We claim that this is a modular form (and hence again a cusp form, as for $\nu = 1$) of weight $k + l + 2\nu$. To see this, first prove the generalization

$$(11) \qquad f^{(\mu)}\Big(\frac{a\tau+b}{c\tau+d}\Big) = \sum_{\lambda=0}^{\mu} \frac{\mu!(k+\mu-1)!}{\lambda!(\mu-\lambda)!(k+\lambda-1)!} c^{\mu-\lambda} (c\tau+d)^{k+\mu+\lambda} f^{(\lambda)}(\tau)$$

of (8) by induction on μ. These transformation formulas can be combined into the single statement that the generating function

$$(12) \qquad \tilde{f}(\tau, X) = \sum_{\mu=0}^{\infty} \frac{1}{\mu!(k+\mu-1)!} f^{(\mu)}(\tau)\, X^\mu \qquad (\tau \in \mathfrak{H}, X \in \mathbb{C})$$

satisfies

$$(13) \qquad \tilde{f}\Big(\frac{a\tau+b}{c\tau+d}, \frac{X}{(c\tau+d)^2}\Big) = (c\tau+d)^k\, e^{cX/(c\tau+d)}\, \tilde{f}(\tau, X) \qquad \Big(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma\Big).$$

From this and the corresponding formula for g it follows that the product

$$\tilde{f}(\tau, X)\, \tilde{g}(\tau, -X) \;=\; \sum_{\nu=0}^{\infty} \frac{(2\pi i)^\nu}{(\nu+k-1)!(\nu+l-1)!} F_\nu(f,g)(\tau)\, X^\nu$$

is multiplied by $(c\tau+d)^{k+l}$ when τ and X are replaced by $\frac{a\tau+b}{c\tau+d}$ and $\frac{X}{(c\tau+d)^2}$, and this proves the modular transformation property of $F_\nu(f,g)$ for every ν. The differential operators F_ν have many applications in the theory of modular forms, some of which will be described in later chapters. As an example, we mention that the generalized theta series with spherical polynomial coefficients $\Theta_{Q,P}$ mentioned in Chapter 0 can be obtained from the simpler theta series Θ_Q by the use of these operators, e.g. if $\theta(\tau) = \sum q^{n^2}$ is the basic theta-series of weight $\frac{1}{2}$ on $\Gamma_0(4)$, then $\frac{2}{3}F_2(\theta, \theta)$ is the function occurring in equation (3) of Chapter 0.

The second way around the problem caused by the extra term in (8) involves the near-modular form G_2 of Section 1.1. Comparing the transformation equations (5) and (8), we find that for any $f \in M_k$ the function $f'(\tau) + 4\pi i k G_2(\tau) f(\tau)$ belongs to M_{k+2}. For instance, for the two forms G_4 and G_6 we have

$$(14)\quad \frac{1}{2\pi i}\, G_4'(\tau) = \frac{7}{10}\, G_6(\tau) - 8\, G_2(\tau)\, G_4(\tau), \qquad \frac{1}{2\pi i}\, G_6'(\tau) = \frac{10}{21}\, G_8(\tau) - 12\, G_2(\tau)\, G_6(\tau)\,.$$

Similarly, by differentiating (3) we find that $G_2'(\tau) + 4\pi i G_2(\tau)^2$ belongs to M_4, whence

$$(15)\qquad \frac{1}{2\pi i}\, G_2'(\tau) = \frac{5}{6}\, G_4(\tau) - 2\, G_2(\tau)^2\,.$$

The formulas (14) and (15) imply that the extension $\mathbb{C}[G_2, G_4, G_6]$ of the ring $M_* = \mathbb{C}[G_4, G_6]$ is closed under differentiation. It follows that if a function f belongs to this ring, then so do the functions f', f'' and f''' and (since the ring has only three generators) these four functions must be algebraically dependent. In particular, G_2 *and all modular forms on* Γ_1 *satisfy third-order differential equations*, e.g., $\frac{i}{\pi}\, G_2''' - 48\, G_2\, G_2'' + 72\, {G_2'}^2 = 0$.

Exercises

1. Show that the series (2) converges absolutely for $k > 2$. (Hint: The number of pairs (m, n) for which $|m\tau + n|$ lies between N and $N+1$ is bounded by a multiple of N, so the series converges like $\sum\limits_{N=1}^{\infty} \dfrac{N}{N^k}$.)

2. Prove that the space M_2 is trivial. (Hint: Show that if $f(\tau) = \sum_{n=0}^{\infty} a(n) q^n$ belongs to M_2 then the integrated function $F(\tau) = a(0)\tau + \sum_{n=1}^{\infty} a(n) q^n / 2\pi i n$ would be Γ_1-invariant up to a constant, i.e., $F(\gamma(\tau)) = F(\tau) + C(\gamma)$ for all $\tau \in \mathfrak{H}$ and $\gamma \in \Gamma_1$ for some constant $C(\gamma)$. The map $C : \Gamma_1 \to \mathbb{C}$ would be a homomorphism and hence 0 since Γ_1 is generated by the elements S and ST of finite order, where S and T are defined as in Exercise 2 of Chapter 0.)

3. Show that a modular form of weight k on Γ_1 vanishes at i if $4 \nmid k$ and at $(\pm 1 + i\sqrt{3})/2$ if $3 \nmid k$ (hint: apply (1) with $\gamma = S$ or ST).

4. Check the various claims made in the final paragraph of §1.2.

5. Check the formula (11) and the other steps in the proof of the modularity of $F_\nu(f, g)$.

The key to the rich internal structure of the theory of modular forms is the existence of a commutative algebra of operators T_n $(n \in \mathbb{N})$ acting on the space M_k of modular forms of weight k. The space M_k has a canonical basis of simultaneous eigenvectors of all the T_n; these special modular forms have the property that their Fourier coefficients $a(n)$ are algebraic integers and satisfy the multiplicative property $a(nm) = a(n)a(m)$ whenever n and m are relatively prime. In particular, their associated Dirichlet series $\sum a(n)n^{-s}$ have Euler products; they also have analytic continuations to the whole complex plane and satisfy functional equations analogous to that of the Riemann zeta function. We will define the operators T_n in Section 2.1 and describe their eigenforms and the associated Dirichlet series in Sections 2.2 and 2.3, respectively. The final section of the chapter describes the modifications of the theory for modular forms on subgroups of $SL_2(\mathbb{Z})$.

2.1. Hecke operators. At the beginning of Chapter 0 we introduced the notion of modular forms of higher weight by giving a bijection

$$\begin{aligned} &F(\Lambda) \mapsto f(\tau) = F(\mathbb{Z}\tau + \mathbb{Z}), \\ &f(\tau) \mapsto F(\Lambda) = \omega_2^{-k} f(\omega_1/\omega_2) \qquad (\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2,\ \Im(\omega_1/\omega_2) > 0) \end{aligned} \tag{1}$$

between functions f in the upper half-plane transforming like modular forms of weight k and functions F of lattices $\Lambda \subset \mathbb{C}$ which are homogeneous of weight $-k$, $F(\lambda\Lambda) = \lambda^{-k}F(\Lambda)$. If we fix a positive integer n, then every lattice Λ has a finite number of sublattices Λ' of index n, and we have an operator T_n on functions of lattices which assigns to such a function F the new function

$$T_n F(\Lambda) = n^{k-1} \sum_{\substack{\Lambda' \subseteq \Lambda \\ [\Lambda:\Lambda']=n}} F(\Lambda') \tag{2}$$

(the factor n^{k-1} is introduced for convenience only). Clearly T_nF is homogeneous of degree $-k$ if F is, so we can transfer the operator to an operator T_n on functions in the upper half-plane which transform like modular forms of weight k. This operator is given explicitly by

$$T_n f(\tau) = n^{k-1} \sum_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1 \backslash \mathcal{M}_n} (c\tau + d)^{-k} f\Big(\frac{a\tau + b}{c\tau + d}\Big) \tag{3}$$

and is called the n**th Hecke operator** in weight k; here $\mathcal{M}_n$ denotes the set of 2×2 integral matrices of determinant n and $\Gamma_1 \backslash \mathcal{M}_n$ the finite set of orbits of $\mathcal{M}_n$ under left multiplication by elements of $\Gamma_1 = SL_2(\mathbb{Z})$. Clearly this definition depends on k and we should more correctly write $T_k(n)f$ or (the standard notation) $f|_kT_n$, but we will consider the weight as fixed and write simply T_nf for convenience. In terms of the "slash operator"

$$(f|_k\gamma)(\tau)=\frac{(ad-bc)^{k/2}}{(c\tau+d)^k}f\big(\frac{a\tau+b}{c\tau+d}\big)\qquad(\gamma=\big(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\big),\ a,b,c,d\in\mathbb{R},\ ad-bc>0)$$

mentioned in Chapter 0, formula (3) can be expressed in the form

$$T_nf(\tau)=n^{\frac{k}{2}-1}\sum_{\mu\in\Gamma_1\backslash\mathcal{M}_n}f|_k\mu.$$

From the fact that $|_k$ is a group operation (i.e. $f|_k(\gamma_1\gamma_2)=(f|_k\gamma_1)|_k\gamma_2$ for γ_1, γ_2 in $GL_2^+(\mathbb{R})$), we see that T_nf is well-defined (changing the orbit representative μ to $\gamma\mu$ with $\gamma\in\Gamma_1$ doesn't affect $f|_k\mu$ because $f|_k\gamma=f$) and again transforms like a modular form of weight k on Γ_1 ($(T_nf)|_k\gamma=T_nf$ for $\gamma\in\Gamma_1$ because $\{\mu\gamma\mid\mu\in\Gamma_1\backslash\mathcal{M}_n\}$ is another set of representatives for $\Gamma_1\backslash\mathcal{M}_n$). Of course, both of these properties are also obvious from the invariant definition (2) and the isomorphism (1).

Formula (3) makes it clear that T_n preserves the property of being holomorphic. We now give a description of the action of T_n on Fourier expansions which shows that T_n also preserves the growth properties at infinity defining modular forms and cusp forms, respectively, and also that the various Hecke operators commute with one another.

Theorem 1. *(i) If $f(\tau)$ is a modular form with the Fourier expansion $\sum_{m=0}^{\infty}a_mq^m$ ($q=e^{2\pi i\tau}$), then the Fourier expansion of T_nf is given by*

$$T_nf(\tau)=\sum_{m=0}^{\infty}\Big(\sum_{d|n,m}d^{k-1}\,a\big(\frac{nm}{d^2}\big)\Big)\,q^m,\tag{4}$$

where $\sum_{d|n,m}$ denotes a sum over the positive common divisors of n and m. In particular, T_nf is again a modular form, and is a cusp form if f is one.
(ii) The Hecke operators in weight k satisfy the multiplication rule

$$T_nT_m=\sum_{d|n,m}d^{k-1}\,T_{nm/d^2}.\tag{5}$$

In particular, $T_nT_m=T_mT_n$ for all n and m and $T_nT_m=T_{nm}$ if n and m are coprime.

Proof. If $\mu=\big(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\big)$ is a matrix of determinant n with $c\neq0$, then we can choose a matrix $\gamma=\big(\begin{smallmatrix}a'&b'\\c'&d'\end{smallmatrix}\big)\in SL_2(\mathbb{Z})$ with $\dfrac{a'}{c'}=\dfrac{a}{c}$, and $\gamma^{-1}\mu$ then has the form $\big(\begin{smallmatrix}*&*\\0&*\end{smallmatrix}\big)$. Hence we can assume that the coset representatives in (3) have the form $\mu=\big(\begin{smallmatrix}a&b\\0&d\end{smallmatrix}\big)$ with $ad=n$, $b\in\mathbb{Z}$. A different choice $\gamma\big(\begin{smallmatrix}a&b\\0&d\end{smallmatrix}\big)$ ($\gamma\in SL_2(\mathbb{Z})$) of representative also has this form if and only if $\gamma=\pm\big(\begin{smallmatrix}1&r\\0&1\end{smallmatrix}\big)$ with $r\in\mathbb{Z}$, in which case $\gamma\big(\begin{smallmatrix}a&b\\0&d\end{smallmatrix}\big)=\pm\big(\begin{smallmatrix}a&b+dr\\0&d\end{smallmatrix}\big)$, so the choice of μ is unique if we require a, $d>0$ and $0\le b<d$. Hence

$$T_nf(\tau)=n^{k-1}\sum_{\substack{a,d>0\\ad=n}}\sum_{b=0}^{d-1}d^{-k}\,f\big(\frac{a\tau+b}{d}\big).$$

Substituting into this the formula $f = \sum a(m)\, q^m$ gives (4) after a short calculation. The second assertion of (i) follows from (4) because all of the exponents of q on the right-hand side are ≥ 0 and the constant term equals $a(0)\sigma_{k-1}(n)$ ($\sigma_{k-1}(n)$ as in 1.1), so vanishes if $a(0) = 0$. The multiplication properties (5) follow from (4) by another easy computation. ■

In the special case when $n = p$ is prime, the formula for the action of T_n reduces to

$$T_p f(\tau) = \frac{1}{p}\sum_{j=0}^{p-1} f\Bigl(\frac{\tau+j}{p}\Bigr) + p^{k-1} f(p\tau) = \sum_{m=0}^{\infty} a(mp)\, q^m + p^{k-1}\sum_{m=0}^{\infty} a(m)\, q^{mp}.$$

The multiplicative property (5) tells us that knowing the T_p is sufficient for knowing all T_n, since if $n > 1$ is divisible by a prime p then $T_n = T_{n/p}T_p$ if $p^2 \nmid n$, $T_n = T_{n/p}T_p - p^{k-1}T_{n/p^2}$ if $p^2|n$.

To end this section, we remark that formula (4), except for the constant term, makes sense also for $n = 0$, the common divisors of 0 and m being simply the divisors of m. Thus the coefficient of q^m on the right is just $a(0)\sigma_{k-1}(m)$ for each $m > 0$. The constant term is formally $a(0)\sum_{d=1}^{\infty} d^{k-1} = a(0)\zeta(1-k)$, but in fact we take it to be $\frac{1}{2}a(0)\zeta(1-k) = -a(0)\dfrac{B_k}{2k}$. Thus we set

$$(6) \qquad T_0 f(\tau) = a(0)\, G_k(\tau) \qquad \Bigl(f = \sum_{m=0}^{\infty} a(m)\, q^m \in M_k\Bigr);$$

in particular, T_0 maps M_k to M_k and $T_0 f = 0$ if f is a cusp form.

2.2. Eigenforms. We have seen that the Hecke operators T_n act as linear operators on the vector space M_k. Suppose that some modular form $f(\tau) = \sum_{m=0}^{\infty} a(m)\, q^m$ is an eigenvector of all the T_n, i.e.,

$$(7) \qquad T_n f = \lambda_n f \qquad (\forall n)$$

for some complex numbers λ_n. This certainly sometimes happens. For instance, if $k = 4$, 6, 8, 10 or 14, then the space M_k is 1-dimensional, spanned by the Eisenstein series G_k, so $T_n G_k$ is necessarily a multiple of G_k for every n. (Actually, we will see in a moment that this is true even if $\dim M_k > 1$.) Similarly, if $k = 12$, 16, 18, 20, 22 or 26, then the space S_k of cusp forms of weight k is 1-dimensional, and since T_n preserves S_k, any element of S_k satisfies (7). From (7) and (4) we obtain the identity

$$(8) \qquad \lambda_n\, a(m) = \sum_{d|n,m} d^{k-1}\, a\bigl(\frac{nm}{d^2}\bigr)$$

by comparing the coefficients of q^m on both sides of (7). In particular, $\lambda_n a(1) = a(n)$ for all n. It follows that $a(1) \neq 0$ if f is not identically zero, so we can normalize f by requiring that $a(1) = 1$. We call a modular form satisfying (7) and the extra condition $a(1) = 1$ a **Hecke form** (the term "normalized Hecke eigenform" is commonly used in the literature). From what we have just said, it follows that a Hecke form has the property

$$\lambda_n = a(n) \qquad (\forall n), \tag{9}$$

i.e., the Fourier coefficients of f are equal to its eigenvalues under the Hecke operators. Equation (5) or (8) now implies the property

$$a(n)\,a(m) = \sum_{d|n,m} d^{k-1}\,a\Bigl(\frac{nm}{d^2}\Bigr) \tag{10}$$

for the coefficients of a Hecke form. In particular, the sequence of Fourier coefficients $\{a(n)\}$ is **multiplicative**, i.e., $a(1) = 1$ and $a(nm) = a(n)a(m)$ whenever n and m are coprime. In particular, $a(p_1^{r_1} \ldots p_l^{r_l}) = a(p_1^{r_1}) \ldots a(p_l^{r_l})$ for distinct primes $p_1, \ldots, p_l$, so the $a(n)$ are determined if we know the values of $a(p^r)$ for all primes p. Moreover, (10) with $n = p^r$, $m = p$ gives the recursion

$$a(p^{r+1}) = a(p)\,a(p^r) - p^{k-1}a(p^{r-1}) \qquad (r \ge 1) \tag{11}$$

for the coefficients $a(p^r)$ for a fixed prime p, so it in fact is enough to know the $a(p)$ (compare the remark following Theorem 1).

Examples. **1.** The form $G_k = -\dfrac{B_k}{2k} + \displaystyle\sum_{m=1}^{\infty} \sigma_{k-1}(m)\,q^m \in M_k$ is a Hecke form for all $k \ge 4$ with $\lambda_n = a(n) = \sigma_{k-1}(n)$ for $n > 0$ and $\lambda_0 = a(0) = -\dfrac{B_k}{2k}$ (cf. (6)). In view of (4), to check this we need only check that the coefficients $a(n)$ of G_k satisfy (10) if n or $m > 0$; this is immediate if n or m equals 0 and can be checked easily for n and m positive by reducing to the case of prime powers (for $n = p^\nu$, $\sigma_{k-1}(n)$ equals $1 + p^{k-1} + \cdots + p^{\nu(k-1)}$, which can be summed as a geometric series) and using the obvious multiplicativity of the numbers $\sigma_{k-1}(n)$.

2. The discriminant function Δ discussed in 1.2 belongs to the 1-dimensional space S_{12} and has 1 as coefficient of q^1, so it is a Hecke form. In particular, (10) holds (with $k = 12$) for the coefficients $a(n)$ of Δ, as we can check for small n using the coefficients given in (7) of 1.2:

$$a(2)a(3) = -24 \times 252 = -6048 = a(6), \quad a(2)^2 = 576 = -1472 + 2048 = a(4) + 2^{11}.$$

This multiplicativity property of the coefficients of Δ was noticed by Ramanujan in 1916 and proved by Mordell a year later by the same argument as we have just given.

The proof that Δ is a simultaneous eigenform of the T_n used the property $\dim S_k = 1$, which is false for $k > 26$. Nevertheless, there exist eigenforms in higher dimensions also; this is Hecke's great discovery. Indeed, we have:

Theorem 2. *The Hecke forms in M_k form a basis of M_k for every k.*

Proof. We have seen that G_k is an eigenform of all T_n. Conversely, any modular form with non-zero constant term which is an eigenform of all T_n $(n \geq 0)$ is a multiple of G_k by virtue of equation (6) of Section 2.1. In view of this and the decomposition $M_k = \langle G_k \rangle \oplus S_k$, it suffices to show that S_k is spanned by Hecke forms and that the Hecke forms in S_k are linearly independent. For this we use the **Petersson scalar product**

$$(f,g) \;=\; \iint_{\mathfrak{H}/\Gamma_1} v^k\, f(\tau)\,\overline{g(\tau)}\,d\mu \qquad (f,g \in S_k), \tag{12}$$

where we have written τ as $u+iv$ and $d\mu$ for the measure $v^{-2}\,du\,dv$ on $\mathfrak{H}$, and the integral is taken over a fundamental domain for the action of Γ_1 on $\mathfrak{H}$. This integral makes sense because the function $v^k|f(\tau)|^2$ and the measure $d\mu$ are both Γ_1-invariant and converges absolutely because $v^k|f(\tau)|^2$ is bounded and $\int_{\mathfrak{H}/\Gamma_1} d\mu < \infty$. It gives S_k the structure of a finite-dimensional Hilbert space. One checks from the definition (3) that the T_n are self-adjoint with respect to this structure, i.e. $(T_n f, g) = (f, T_n g)$ for all f, $g \in S_k$ and $n > 0$. (For $n = 0$, of course, T_n is the zero operator on S_k by equation (6).) Also, the T_n commute with one another, as we have seen. A well-known theorem of linear algebra then asserts that S_k is spanned by simultaneous eigenvectors of all the transformations T_n, and we have already seen that each such eigenform is uniquely expressible as a multiple of a Hecke form satisfying (10). Moreover, for a Hecke form we have

$$\begin{aligned} a(n)\,(f,f) &= (a(n)f,f) = (\lambda_n f,f) = (T_n f,f) \\ &= (f,T_n f) = (f,\lambda_n f) = (f,a(n)f) = \overline{a(n)}\,(f,f) \end{aligned}$$

by the self-adjointness of T_n and the sesquilinearity of the scalar product. Therefore the Fourier coefficients of f are real. If $g = \sum b(n)\,q^n$ is a second Hecke form in S_k, then the same computation shows that

$$a(n)\,(f,g) = (T_n f,g) = (f,T_n g) = \overline{b(n)}\,(f,g) = b(n)\,(f,g)$$

and hence that $(f,g) = 0$ if $f \neq g$. Thus the various Hecke forms in S_k are mutually orthogonal and *a fortiori* linearly independent. ■

We also have

Theorem 3. *The Fourier coefficients of a Hecke form $f \in S_k$ are real algebraic integers of degree $\leq \dim S_k$.*

Proof. The space S_k is spanned by forms all of whose Fourier coefficients are integral (this follows easily from the discussion in Section 1.2. By formula (4), the lattice L_k of all such forms is mapped to itself by all T_n. Let $f_1, \dots, f_d$ $(d = \dim_{\mathbb{C}} S_k = \mathrm{rk}_{\mathbb{Z}} L_k)$ be a basis for L_k over $\mathbb{Z}$. Then the action of T_n with respect to this basis is given by a $d \times d$ matrix with coefficients in $\mathbb{Z}$, so the eigenvalues of T_n are algebraic integers of degree $\le d$. By (9), these eigenvalues are precisely the Fourier coefficients of the d Hecke forms in S_k. That the coefficients of Hecke forms are real was already checked in proving Theorem 2. ■

From the proof of the theorem, we see that the trace of T_n $(n > 0)$ acting on M_k or S_k is the trace of a $(d+1)\times(d+1)$ or $d\times d$ matrix with integral coefficients and hence is an integer. This trace is given in closed form by the Eichler-Selberg trace formula, which will be discussed in Chapter 8.

Example. The space S_{24} is 2-dimensional, spanned by

$$\Delta(\tau)^2 = 0q + q^2 - 48\,q^3 + 1080\,q^4 + \cdots$$

and

$$(240G_4(\tau))^3\,\Delta(\tau) = q + 696\,q^2 + 162252\,q^3 + 12831808\,q^4 + \cdots$$

If $f \in S_{24}$ is a Hecke form, then f must have the form $(240G_4)^3\Delta + \lambda\Delta^2$ for some $\lambda \in \mathbb{C}$, since the coefficient of q^1 must be 1. Hence its second and fourth coefficients are given by

$$a(2) = 696 + \lambda, \qquad a(4) = 12831808 + 1080\,\lambda.$$

The property $a(2)^2 = a(4) + 2^{23}$ $(n = m = 2$ in (10)) now leads to the quadratic equation

$$\lambda^2 + 312\,\lambda - 20736000 = 0$$

for λ. Hence any Hecke form in S_{24} must be one of the two functions

$$f_1,\ f_2 \;=\; (240G_4)^3\,\Delta + \bigl(-156 \pm 12\sqrt{144169}\bigr)\,\Delta^2.$$

Since Theorem 2 says that S_{24} must contain exactly two Hecke forms, f_1 and f_2 are indeed eigenvectors with respect to all the T_n. This means, for example, that we would have obtained the same quadratic equation for λ if we had used the relation $a(2)a(3) = a(6)$ instead of $a(2)^2 = a(4) + 2^{23}$. The coefficients $a_1(n)$, $a_2(n)$ of f_1 and f_2 are conjugate algebraic integers in the real quadratic field $\mathbb{Q}(\sqrt{144169})$.

2.3. L-series. The natural reflex of a number-theorist confronted with a multiplicative function $n \mapsto a(n)$ is to form the Dirichlet series $\sum\limits_{n=1}^{\infty} a(n)\,n^{-s}$, the point being that the multiplicative property implies that $a(p_1^{r_1}\dots p_l^{r_l}) = a(p_1^{r_1})\dots a(p_l^{r_l})$ and hence that this Dirichlet series has an Euler product $\prod\limits_{p\text{ prime}} \bigl(\sum\limits_{r\ge 0} a(p^r)\,p^{-rs}\bigr)$. We therefore define the **Hecke L-series** of a modular form $f(\tau) = \sum_{m=0}^{\infty} a(m)\,q^m \in M_k$ by

$$L(f,s) = \sum_{m=1}^{\infty} \frac{a(m)}{m^s} \tag{13}$$

(notice that we have ignored $a(0)$ in this definition; what else could we do?). Thus if f is a Hecke form we have an Euler product

$$L(f,s) = \prod_{p \text{ prime}} \left(1 + \frac{a(p)}{p^s} + \frac{a(p^2)}{p^{2s}} + \cdots\right)$$

because the coefficients $a(m)$ are multiplicative. But in fact we can go further, because the recursion (11) implies that for each prime p the generating function $A_p(x) = \sum a(p^r)x^r$ satisfies

$$\begin{aligned} A_p(x) = 1 + \sum_{r=0}^{\infty} a(p^{r+1})\,x^{r+1} &= 1 + \sum_{r=0}^{\infty} a(p)\,a(p^r)\,x^{r+1} - \sum_{r=1}^{\infty} p^{k-1}\,a(p^{r-1})\,x^{r+1} \\ &= 1 + a(p)\,x\,A_p(x) - p^{k-1}\,x^2\,A_p(x) \end{aligned}$$

and hence that

$$A_p(x) = \frac{1}{1 - a(p)x + p^{k-1}x^2}.$$

Therefore, replacing x by p^{-s} and multiplying over all primes p, we find finally

(14) $$L(f,s) = \prod_p \frac{1}{1 - a(p)p^{-s} + p^{k-1-2s}} \qquad (f \in M_k \text{ a Hecke form}).$$

Examples. **1.** For $f = G_k$ we have

$$\begin{aligned} a(p^r) &= 1 + p^{k-1} + \cdots + p^{r(k-1)} = \frac{p^{(r+1)(k-1)} - 1}{p^{k-1} - 1}, \\ A_p(x) &= \sum_{r=0}^{\infty} \frac{p^{(r+1)(k-1)} - 1}{p^{k-1} - 1}\,x^r = \frac{1}{\bigl(1 - p^{k-1}x\bigr)\bigl(1 - x\bigr)} \\ L(G_k, s) &= \prod_p \frac{1}{1 - \sigma_{k-1}(p)p^{-s} + p^{k-1-2s}} = \prod_p \frac{1}{(1 - p^{k-1-s})(1 - p^{-s})} \\ &= \zeta(s-k+1)\zeta(s), \end{aligned}$$

where $\zeta(s)$ is the Riemann zeta function. (Of course, we could see this directly: the coefficient of n^{-s} in $\zeta(s-k+1)\zeta(s) = \sum_{d,e \geq 1} \frac{d^{k-1}}{(de)^s}$ is clearly $\sigma_{k-1}(n)$ for each $n \geq 1$.)

2. For $f = \Delta$ we have

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

where $\tau(n)$, the Ramanujan tau-function, denotes the coefficient of q^n in Δ; this identity summarizes all the multiplicative properties of $\tau(n)$ discovered by Ramanujan.

Of course, the Hecke L-series would be of no interest if their definition were merely formal. However, these series converge in a half-plane and define functions with nice analytic properties, as we now show.

Theorem 4. *(i) The Fourier coefficients $a(m)$ of a modular form of weight k satisfy the growth estimates*

$$a(n) = O\big(n^{k-1}\big) \qquad (f \in M_k), \qquad\qquad a(n) = O\big(n^{\frac{k}{2}}\big) \qquad (f \in S_k). \tag{15}$$

Hence the L-series $L(f,s)$ converges absolutely and locally uniformly in the half-plane $\Re(s) > k$, and in the larger half-plane $\Re(s) > \dfrac{k}{2} + 1$ if f is a cusp form.

(ii) $L(f,s)$ has a meromorphic continuation to the whole complex plane. It is holomorphic everywhere if f is a cusp form and has exactly one singularity, a simple pole of residue $\dfrac{(2\pi i)^k}{(k-1)!}\,a(0)$ at $s = k$, otherwise. The meromorphically extended function satisfies the functional equation

$$(2\pi)^{-s}\,\Gamma(s)\,L(f,s) \;=\; (-1)^{\frac{k}{2}}(2\pi)^{s-k}\,\Gamma(k-s)\,L(f,k-s).$$

Proof. (i) Since the estimate $a(n) = \mathrm{O}(n^{k-1})$ is obvious for the Eisenstein series G_k (we have $\sigma_{k-1}(n) = n^{k-1}\sum_{d|n} d^{-k+1} < n^{k-1}\sum_{d=1}^{\infty} d^{-k+1} = \zeta(k-1)\,n^{k-1}$), and since every modular form of weight k is a combination of G_k and a cusp form, we need only prove the second estimate in (15). If f is a cusp form then by definition we have $|f(\tau)| < Mv^{-k/2}$ for some constant $M > 0$ and all $\tau = u + iv \in \mathfrak{H}$. On the other hand, for any $n \geq 1$ and $v > 0$ we have

$$a(n) = \int_0^1 f(u+iv)\,e^{-2\pi i n(u+iv)}\,du.$$

Hence

$$|a(n)| \leq M\,v^{-k/2}\,e^{2\pi n v},$$

and choosing $v = 1/n$ gives the desired conclusion. (This argument, like most of the rest of this chapter, is due to Hecke.)

(ii) This follows immediately from the "functional equation principle" in Appendix A2, since the function

$$\phi(v) \;=\; f(iv) - a(0) = \sum_{n=1}^{\infty} a(n)\,e^{-2\pi n v} \qquad\qquad (v > 0)$$

is exponentially small at infinity and satisfies the functional equation

$$\phi\big(\frac{1}{v}\big) = f\big(\frac{-1}{iv}\big) - a(0) \;=\; (iv)^k\,f(iv) - a(0) \;=\; (-1)^{\frac{k}{2}}v^k\,\phi(v) + (-1)^{\frac{k}{2}}a(0)\,v^k - a(0)$$

and its Mellin transform $\int_0^\infty \phi(v)\,v^{s-1}\,dv$ equals $(2\pi)^{-s}\Gamma(s)L(f,s)$. ■

The first estimate in (15) is clearly best possible, but the second one can be improved. The estimate $a(n) = \mathrm{O}(n^{\frac{k}{2}-\frac{1}{5}+\epsilon})$ for the Fourier coefficients of cusp forms on Γ_1 was found by Rankin in 1939 as an application of the Rankin-Selberg method which will be discussed in Chapter 6. This was later improved to $a(n) = \mathrm{O}(n^{\frac{k}{2}-\frac{1}{4}+\epsilon})$ by Selberg as an application of Weil's estimates of Kloosterman sums. The estimate

$$(16) \qquad a(n) = \mathrm{O}(n^{\frac{k-1}{2}+\epsilon}) \qquad \left(f = \sum a(n)\, q^n \,\in\, S_k\right),$$

conjectured by Ramanujan for $f = \Delta$ in 1916 and by Petersson in the general case, remained an open problem for many years. It was shown by Deligne in 1969 to be a consequence of the Weil conjectures on the eigenvalues of the Frobenius operator in the l-adic cohomology of algebraic varieties in positive characteristic; five years later he proved the Weil conjectures, thus establishing (16). Using the form of the generating function $A_p(x)$ given above, one sees that (16) is equivalent to

$$(17) \qquad |a(p)| \le 2p^{(k-1)/2} \qquad (p \text{ prime}).$$

In particular, for the Ramanujan tau-function $\tau(n)$ (coefficient of q^n in Δ) one has

$$(18) \qquad |\tau(p)| \le 2p^{11/2} \qquad (p \text{ prime}).$$

Deligne's proof of (18) uses the full force of Grothendieck's work in algebraic geometry and its length, if written out from scratch, has been estimated at 2000 pages; in his book on mathematics and physics, Manin cites this as a probable record for the ratio "length of proof : length of statement" in the whole of mathematics.

2.4. Forms of higher level. In these notes, we usually restrict attention to the full modular group $\Gamma_1 = SL_2(\mathbb{Z})$ rather than subgroups because most aspects of the theory can be seen there. However, in the case of the theory of Hecke operators there are some important differences, which we now describe. We will discuss only the subgroups $\Gamma_0(N) = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1 \mid c \equiv 0 \pmod{N}\}$ which were already introduced in Chapter 0.

First of all, the definition of T_n must be modified. In formula (3) we must replace Γ_1 by $\Gamma = \Gamma_0(N)$ and $\mathcal{M}_n$ by the set of integral matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ of determinant n satisfying $c \equiv 0 \pmod{N}$ and $(a, N) = 1$. Again the coset representatives of $\Gamma \backslash \mathcal{M}_n$ can be chosen to be upper triangular, but the extra condition $(a, N) = 1$ means that we have fewer representatives than before if $(n, N) > 1$. In particular, for p a prime dividing N we have $T_p f(\tau) = \sum a(mp) q^m$ and $T_{p^r} = (T_p)^r$ rather than the more complicated formulas given in the remark following Theorem 1 in 2.1; more generally, for n arbitrary the operation of T_n is given by the same formula (4) as before but with the extra condition $(d, N) = 1$ added to the inner sum, and similarly for the multiplicativity relation (5).

The other main difference with the case $N = 1$ comes from the existence of so-called "old forms." If N' is a proper divisor of N, then $\Gamma_0(N)$ is a subgroup of $\Gamma_0(N')$ and every modular form $f(\tau)$ of weight k on $\Gamma_0(N')$ is *a fortiori* a modular form on $\Gamma_0(N)$. More generally, $f(M\tau)$ is a modular form of weight k on $\Gamma_0(N)$ for each positive divisor M of

N/N', since

$$\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N) \;\Rightarrow\; \left(\begin{smallmatrix} a & bM \\ c/M & d \end{smallmatrix}\right) \in \Gamma_0(N')$$
$$\Rightarrow\; f\bigl(M\,\frac{a\tau+b}{c\tau+d}\bigr) = f\bigl(\frac{a(M\tau)+bM}{(c/M)(M\tau)+d}\bigr) = (c\tau+d)^k f(M\tau).$$

The subspace of $M_k(\Gamma_0(N))$ spanned by all forms $f(M\tau)$ with $f \in M_k(\Gamma_0(N'))$, $MN'|N$, $N' \neq N$, is called the space of **old forms**. (This definition must be modified slightly if $k=2$ to include also the modular forms $\sum_{M|N} c_M G_2^*(M\tau)$ with $c_M \in \mathbb{C}$, $\sum_{M|N} M^{-1}c_M = 0$, where G_2^* is the non-holomorphic Eisenstein series of weight 2 on Γ_1 introduced in 1.1, as old forms, even though G_2^* itself is not in $M_2(\Gamma_1)$.) Since the old forms can be considered by induction on N as already known, one is interested only in the "rest" of $M_k(\Gamma_0(N))$. The answer here is quite satisfactory: $M_k(\Gamma_0(N))$ has a canonical splitting as the direct sum of the subspace $M_k(\Gamma_0(N))^{\mathrm{old}}$ of old forms and a certainly complementary space $M_k(\Gamma_0(N))^{\mathrm{new}}$ (for cusp forms, $S_k(\Gamma_0(N))^{\mathrm{new}}$ is just the orthogonal complement of $S_k(\Gamma_0(N))^{\mathrm{old}}$ with respect to the Petersson scalar product), and if we define a **Hecke form of level** N to be a form in $M_k(\Gamma_0(N))^{\mathrm{new}}$ which is an eigenvector of T_n for all n prime to N and with $a(1)=1$, then the Hecke forms are in fact eigenvectors of all the T_n, they form a basis of $M_k(\Gamma_0(N))^{\mathrm{new}}$, and their Fourier coefficients are real algebraic integers as before. For the pth Fourier coefficient (p prime) of a Hecke form in $S_k(\Gamma_0(N))^{\mathrm{new}}$ we have the same estimate (15) as before if $p \nmid N$, while the eigenvalue with respect to T_p when $p|N$ equals 0 if $p^2|N$ and $\pm p^{(k-2)/2}$ otherwise. Finally, there is no overlapping between the newforms of different level or between the different lifts $f(M\tau)$ of forms of the same level, so that we have a canonical direct sum decomposition

$$M_k(\Gamma_0(N)) = \bigoplus_{MN'|N} \bigl\{ f(M\tau) \mid f \in M_k(\Gamma_0(N'))^{\mathrm{new}} \bigr\}$$

and a canonical basis of $M_k(\Gamma_0(N))$ consisting of the functions $f(M\tau)$ where $M|N$ and f is a Hecke form of some level N' dividing N/M.

As already stated, the Fourier coefficients of Hecke forms of higher level are real algebraic integers, just as before. However, there is a difference with the case $N=1$. For forms of level 1, Theorem 3 apparently always is sharp: in all cases which have been calculated, the number field generated by the Fourier coefficients of a Hecke cusp form of weight k has degree equal to the full dimension d of the space S_k, which is then spanned by a single form and its algebraic conjugates (cf. the example $k=24$ given above). For forms of higher level, on the other hand, there are in general further splittings. The general situation is that $S_k(\Gamma_0(N))^{\mathrm{new}}$ splits as the sum of subspaces of some dimensions $d_1, \dots, d_r \geq 1$, each of which is spanned by some Hecke form, with Fourier coefficients in a totally real number field K_i of degree d_i over $\mathbb{Q}$, and the algebraic conjugates of this form (i.e. the forms obtained by considering the various embeddings $K_i \hookrightarrow \mathbb{R}$). In general the number r and the dimensions d_i are unknown; the known theory implies certain necessary splittings of $S_k(\Gamma_0(N))^{\text{new}}$, but there are often further splittings which we do not know how to predict.

Examples. **1.** $k=2$, $N=11$. Here $\dim M_k(\Gamma_0(N))=2$. As well as one old form, the Eisenstein series

$$G_2^*(\tau)-11G_2^*(11\tau)=\frac{5}{12}+\sum_{n=1}^{\infty}\Bigl(\sum_{\substack{d\mid n\\ 11\nmid d}} d\Bigr)\,q^n$$

of weight 2, there is one new form

$$f(\tau)=\sqrt[12]{\Delta(\tau)\Delta(11\tau)}=q-2q^2-q^3+2q^4+q^5+2q^6-2q^7+\cdots,$$

with Fourier coefficients in $\mathbb{Z}$. This form corresponds as in the Taniyama-Weil conjecture mentioned in Chapter 0 to the elliptic curve $y^2-y=x^3-x^2$, i.e., the number of solutions of $y^2-y=x^3-x^2$ in integers modulo p is given by $p-a(p)$ for every prime p.

2. $k=2$, $N=23$. Again $M_k(\Gamma_0(N))^{\text{old}}$ is 1-dimensional, spanned by $G_2^*(\tau)-NG_2^*(N\tau)$, but this time $M_k(\Gamma_0(N))^{\text{new}}=S_k(\Gamma_0(N))^{\text{new}}$ is 2-dimensional, spanned by the Hecke form

$$f_1=q-\frac{1-\sqrt5}{2}\,q^2+\sqrt5\,q^3-\frac{1+\sqrt5}{2}\,q^4-(1-\sqrt5)\,q^5-\frac{5-\sqrt5}{2}\,q^6+\cdots$$

with coefficients in $\mathbb{Z}+\mathbb{Z}\,\dfrac{1+\sqrt5}{2}$ and the conjugate form

$$f_2=q-\frac{1+\sqrt5}{2}\,q^2-\sqrt5\,q^3-\frac{1-\sqrt5}{2}\,q^4-(1+\sqrt5)\,q^5-\frac{5+\sqrt5}{2}\,q^6+\cdots$$

obtained by replacing $\sqrt5$ by $-\sqrt5$ everywhere in f_1.

3. $k=2$, $N=37$. Again $M_k(\Gamma_0(N))^{\text{old}}$ is spanned by $G_2^*(\tau)-NG_2^*(N\tau)$ and $M_k(\Gamma_0(N))^{\text{new}}=S_k(\Gamma_0(N))^{\text{new}}$ is 2-dimensional, but this time the two Hecke forms of level N

$$f_1=q-2q^2-3q^3+2q^4-2q^5+6q^6-q^7+\cdots$$

and

$$f_2=q+0q^2+q^3-2q^4+0q^5+0q^6-q^7+\cdots$$

both have coefficients in $\mathbb{Z}$. These forms correspond à la Taniyama-Weil to the elliptic curves $y^2-y=x^3-x$ and $y^2-y=x^3+x^2-3x+1$, respectively.

4. $k=4$, $N=13$. Here $M_k(\Gamma_0(N))^{\text{old}}$ is spanned by the two Eisenstein series $G_4(\tau)$ and $G_4(N\tau)$ and the space $M_k(\Gamma_0(N))^{\text{new}}=S_k(\Gamma_0(N))^{\text{new}}=S_k(\Gamma_0(N))$ is 3-dimensional, spanned by the forms

$$f_1,\,f_2\;=\;q+\frac{1\pm\sqrt{17}}{2}\,q^2+\frac{5\mp3\sqrt{17}}{2}\,q^3-\frac{7\mp\sqrt{17}}{2}\,q^4+\cdots$$

with coefficients in the real quadratic field $\mathbb{Q}(\sqrt{17})$ and the form

$$f_3 = q - 5q^2 - 7q^3 + 17q^4 - 7q^5 + 35q^6 - 13q^7 - \cdots$$

with coefficients in $\mathbb{Q}$.

Finally, there are some differences between the L-series in level 1 and in higher level. First of all, the form of the Euler product for the L-series of a Hecke form (eq. (14)) must be modified slightly: it is now

$$L(f,s) = \prod_{p\nmid N} \frac{1}{1 - a(p)p^{-s} + p^{k-1-2s}} \prod_{p|N} \frac{1}{1 - a(p)p^{-s}}. \tag{19}$$

More important, $L(f,s)$, although it converges absolutely in the same half-plane as before and again has a meromorphic continuation with at most a simple pole at $s = k$, in general *does not have a functional equation for every* $f \in M_k(\Gamma_0(N))$, because we no longer have the element $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \in \Gamma$ to force the symmetry of $f(iv)$ with respect to $v \mapsto \dfrac{1}{v}$. Instead, we have the **Fricke involution**

$$w_N : f(\tau) \mapsto w_N f(\tau) = N^{-\frac{k}{2}}\, \tau^{-k}\, f\Big(\frac{-1}{N\tau}\Big)$$

which acts on the space of modular forms of weight k on $\Gamma_0(N)$ because the element $\left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right)$ of $GL_2^+(\mathbb{R})$ normalizes the group $\Gamma_0(N)$. This involution splits $M_k(\Gamma_0(N))$ into the direct sum of two eigenspaces $M_k^\pm(\Gamma_0(N))$, and if f belongs to $M_k^\pm(\Gamma_0(N))$ then

$$(2\pi)^{-s}\, N^{s/2}\, \Gamma(s)\, L(f,s) = \pm(-1)^{k/2}\, (2\pi)^{s-k}\, N^{(k-s)/2}\, \Gamma(k-s)\, L(f,k-s).$$

(For $N = 1$ we have $w_N \equiv \mathrm{Id}$ since $\left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right) \in \Gamma_0(N)$ in this case, so $M_k^- = \{0\}$ for all k, but for all other values of N the dimension of $M_k^+(\Gamma_0(N))$ is asympotically one-half the dimension of $M_k(\Gamma_0(N))$ as $k \to \infty$.) The involution w_N preserves the space $M_k(\Gamma_0(N))^{\mathrm{new}}$ and commutes with all Hecke operators T_n there (whereas on the full space $M_k(\Gamma_0(N))$ it commutes with T_n only for $(n,N) = 1$). In particular, each Hecke form of level N is an eigenvector of w_N and therefore has an L-series satisfying a functional equation. In our example **3** above, for instance, the Eisenstein series $G_2^*(\tau) - 37G_2^*(37\tau)$ and the cusp form f_2 are anti-invariant under w_{37} and therefore have plus-signs in the functional equations of their L-series, while f_1 is invariant under w_{37} and has an L-series with a minus sign in its functional equation. In particular, the L-series of f_1 vanishes at $s = 1$, which is related by the famous Birch-Swinnerton-Dyer conjecture to the fact that the equation of the corresponding elliptic curve $y^2 - y = x^3 - x$ has an infinite number of rational solutions.

Exercises

1. Using (4), verify the multiplicative property (5) of the Hecke operators.
2. Verify that the coefficients $a(n) = \sigma_{k-1}(n)$ of G_k satisfy the identities (11) and (10).

The basic statement is that, given an r-dimensional lattice in which the length-squared of any vector is an integer, the multiplicities of these lengths are the Fourier coefficients of a modular form of weight $\frac{r}{2}$. By choosing a basis of the lattice, we can think of it as the standard lattice $\mathbb{Z}^r \subset \mathbb{R}^r$; the square-of-the-length function then becomes a quadratic form Q on $\mathbb{R}^r$ which assumes integral values on $\mathbb{Z}^r$, and the modular form in question is the **theta series**

$$\Theta_Q(\tau) = \sum_{x\in\mathbb{Z}^r} q^{Q(x)} = \sum_{n=0}^{\infty} r_Q(n)\, q^n\,, \qquad r_Q(n) = \#\{x\in\mathbb{Z}^r \mid Q(x)=n\}\,.$$

In general this will not be a modular form on the full modular group $\Gamma_1 = SL_2(\mathbb{Z})$, but on a subgroup of finite index. Examples and the exact transformation law of Θ_Q will be given in 3.1, after which we will discuss generalizations and special cases of these series.

3.1. Theta series of definite quadratic forms. As a first example, let $r=2$ and Q be the modular form $Q(x_1,x_2) = x_1^2 + x_2^2$, so that the associated theta-series, whose Fourier development begins

$$\Theta_Q(\tau) = 1 + 4q + 4q^2 + 0q^3 + 4q^4 + 8q^5 + 0q^6 + 0q^7 + 4q^8 + \cdots\,,$$

counts the number of representations of integers as sums of two squares. This is a modular form of weight 1, not on Γ_1 (for which, as we have seen, there are no modular forms of odd weight), but on the subgroup $\Gamma_0(4)$ consisting of matrices $\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right)$ with c divisible by 4; specifically, we have

$$\Theta_Q\left(\frac{a\tau+b}{c\tau+d}\right) = (-1)^{\frac{d-1}{2}}(c\tau+d)\,\Theta_Q(\tau) \tag{1}$$

for all $\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right) \in \Gamma_0(4)$. To prove this, we observe that Θ_Q is the square of the Jacobi theta function $\theta(\tau) = \sum_{n\in\mathbb{Z}} q^{n^2}$ introduced in Chapter 0. The identity

$$\sum_{n=-\infty}^{\infty} e^{-\pi a n^2} = a^{-\frac{1}{2}} \sum_{n=-\infty}^{\infty} e^{-\pi a^{-1} n^2}$$

proved in Appendix A1 as a consequence of the Poisson summation formula says that $\theta(-1/4\tau) = \sqrt{2\tau/i}\,\theta(\tau)$ and hence $\Theta_Q(-1/4\tau) = (2\tau/i)\,\Theta_Q(\tau)$ for τ on the positive imaginary axis, but then this identity holds for all $\tau \in \mathfrak{H}$ by analytic continuation. Combining this with the trivial relation $\Theta_Q(\tau+1) = \Theta_Q(\tau)$ we find

$$\begin{aligned}\Theta_Q\big(\frac{\tau}{4\tau+1}\big) &= \Theta_Q\left(\frac{-1}{4\big(-\dfrac{1}{4\tau}-1\big)}\right) = \frac{2}{i}\big(-\frac{1}{4\tau}-1\big)\,\Theta_Q\big(-\frac{1}{4\tau}-1\big)\\ &= \big(\frac{4\tau+1}{2\tau/i}\big)\,\Theta_Q\big(-\frac{1}{4\tau}\big) = (4\tau+1)\,\Theta_Q(\tau)\,.\end{aligned}$$

Hence (1) holds for the two matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 0 \\ 4 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Since these two matrices (and $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$) generate $\Gamma_0(4)$ and equation (1) is compatible with multiplication (Exercises 1 and 2), this proves (1). Alternatively, one can directly use the transformation laws of $\theta(\tau)$ under the transformations $\tau \mapsto \tau + 1$ and $\tau \mapsto -1/4\tau$ to obtain the transformation behaviour of Θ_Q under the group they generate, and then observe that this group contains $\Gamma_0(4)$ as a subgroup (Exercises 3 and 4).

Equation (1) describes a modular form of a type we have not seen before, namely one with character. If N is a natural number and χ a Dirichlet character modulo N (cf. Appendix A3), then a **modular form of weight k, level N and character χ** is a holomorphic function f on $\mathfrak{H}$ satisfying the transformation law

$$(2) \qquad f\Big(\frac{a\tau+b}{c\tau+d}\Big) \;=\; \chi(d)\,(c\tau+d)^k\,f(\tau) \qquad\qquad \forall \tau \in \mathfrak{H},\ \Big(\begin{matrix} a & b \\ c & d \end{matrix}\Big) \in \Gamma_0(N)$$

as well as the usual growth conditions. The space of such forms will be denoted by $M_k(\Gamma_0(N), \chi)$ and the subspace of cusp forms by $S_k(\Gamma_0(N), \chi)$. Applying (2) to the matrix $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, which belongs to $\Gamma_0(N)$ for every N, we see that $M_k(\Gamma_0(N), \chi) = \{0\}$ unless $\chi(-1) = (-1)^k$, i.e., unless the weight k and character χ have the same parity (both even or both odd). Note also that the factor $\chi(d)$ in (2) can be omitted when $\chi = \chi_0$ is the principal character modulo N, since $ad - bc = 1$ and $N|c$ imply that d is prime to N, so $M_k(\Gamma_0(N), \chi_0)$ is simply $M_k(\Gamma_0(N))$.

Equation (1) tells us that Θ_Q for $Q(x_1, x_2) = x_1^2 + x_2^2$ belongs to $M_1(\Gamma_0(4), \chi_4)$ where χ_4 is the (unique) odd Dirichlet character modulo 4. This space can be shown to be one-dimensional, generated by the Eisenstein series

$$G_{1,\chi_4}(\tau) = \frac{1}{4} + \sum_{n=1}^{\infty} \Big(\sum_{d|n} \chi_4(n) \Big)\, q^n \;.$$

It follows by comparing constant terms that $\Theta_Q = 4G_{1,\chi_4}$ and hence that $r_2(n)$, the number of representations of n as a sum of two squares, equals $4\sum_{d|n}\chi_4(d)$ for every $n > 0$, a classical theorem of number theory which can be interpreted in terms of lthe arithmetic of the Gaussian integers $\mathbb{Z}[i]$. Similarly, the fact that $\Theta_Q^2 = \theta^4$ and $\Theta_Q^4 = \theta^8$ belong to the spaces $M_2(\Gamma_0(4))$ and $M_4(\Gamma_0(4))$ with bases $\{G_2(\tau) - 2G_2(2\tau),\, G_2(\tau) - 4G_2(4\tau)\}$ and $\{G_4(\tau),\, G_4(2\tau),\, G_4(4\tau)\}$, respectively, gives simple closed formulas for $r_4(n)$ and $r_8(n)$, as already described in Chapter 0.

The general theorem says that says that Θ_Q for any quadratic form $Q : \mathbb{Z}^r \to \mathbb{Z}$ in an even number of variables (the case r odd will be discussed later when we treat modular forms of half-integral weight) is a modular form of weight $r/2$ and some level and character, which are determined as follows. We can write the quadratic form $Q(x)$ as $\sum_{i,j} c_{ij} x_i x_j$ with $c_{ij} \in \mathbb{Z}$ or more compactly as $x^t C x$ where $C = (c_{ij})$ is an $r \times r$ matrix with integer coefficients. It is more convenient to replace C by $A = C + C^t = (a_{ij})$, $a_{ij} = c_{ij} + c_{ji}$; then A is an **even symmetric matrix** (i.e. $a_{ij} = a_{ji}$, $a_{ij} \in \mathbb{Z}$, $a_{ii} \in 2\mathbb{Z}$ for all i, j)

and $Q(x) = \frac{1}{2}x^t Ax$. The **level** of Q is by definition the smallest positive integer N such that the matrix $A^* = N\,A^{-1}$ is again an even matrix. Since $\det(A)\,A^{-1}$ can be checked (using Exercise 5) to be even, we have $N|\det(A)$; on the other hand, the determinant $N^r\det(A)^{-1}$ of A^* is integral, so $\det(A)|N^r$. Hence N and $\det(A)$ have the same prime factors. Moreover, the **discriminant** $D = (-1)^{r/2}\det(A)$ of Q is congruent to 0 or 1 modulo 4 (Exercise 5), so the Kronecker symbol $\left(\frac{D}{\cdot}\right)$ is well-defined (cf. Appendix A3).

Theorem. *Let $Q : \mathbb{Z}^r \to \mathbb{Z}$ be a positive definite quadratic form in r variables, r even. Let N be the level and D the discriminant of Q. Then Θ_Q is a modular form of weight $r/2$, level N and character $\left(\frac{D}{\cdot}\right)$.*

We sketch the proof only in the special case $N = 1$, i.e. of the following result:

Theorem. *Let $Q(x) = \frac{1}{2}x^t Ax$ where A is a positive definite unimodular ($\det(A) = 1$) even symmetric matrix. Then r is divisible by 8 and Θ_Q belongs to $M_{r/2}(SL_2(\mathbb{Z}))$.*

Proof. The Poisson summation formula as discussed in Appendix A1 generalizes immediately to $\sum_{x\in\mathbb{Z}^r} f(x) = \sum_{x\in\mathbb{Z}^r} \tilde{f}(x)$, where $f : \mathbb{R}^r \to \mathbb{C}$ is any smooth function which decays rapidly for $|x|$ large and $\tilde{f}(x) = \int_{\mathbb{R}^r} f(y)e^{2\pi i x\cdot y}dy$ denotes the r-dimensional Fourier transform of f. By Exercise 6, the Fourier transform of $f(x) = e^{-2\pi t^{-1}Q(x)}$ ($t > 0$) is $t^{r/2}\,e^{-2\pi t Q^*(x)}$, where $Q^*(x) = \frac{1}{2}x^t A^* x$ is the quadratic form associated to $A^* = A^{-1}$. Hence $\Theta_Q(i/t) = t^{r/2}\,\Theta_{Q^*}(it)$. Moreover, Θ_Q and Θ_{Q^*} are the same function, as one sees by making the change of variables $x \mapsto Ax$ (which maps $\mathbb{Z}^r$ isomorphically onto $\mathbb{Z}^r$), so $\Theta_Q(-1/\tau) = (\tau/i)^{r/2}\Theta_Q(\tau)$ for τ on the positive imaginary axis and hence, by analytic continuation, for all τ in the upper half-plane. But $\Theta_Q(\tau)$ is obviously invariant under $\tau \mapsto \tau+1$, so $\Theta_Q(1-1/\tau) = (\tau/i)^{r/2}\Theta_Q(\tau)$. The fact that $\tau \mapsto 1-1/\tau$ is a transformation of order 3 now implies that $i^{-3r/2} = 1$ and hence that r is divisible by 8; that Θ_Q is a modular form of weight $r/2$ on $\Gamma_1 = SL_2(\mathbb{Z})$ then follows because Γ_1 is generated by $\tau \mapsto -1/\tau$ and $\tau \mapsto \tau+1$. ■

The proof of the theorem for general Q is harder because, although we can still use the Poisson summation formula to get $\Theta_Q\left(-\frac{1}{N\tau}\right) = \frac{(N\tau/i)^{r/2}}{\det(A)^{1/2}}\,\Theta_{Q^*}(\tau)$, the transformations $\tau \mapsto \tau+1$ and $\tau \mapsto -1/N\tau$ no longer generate $\Gamma_0(N)$ (or, as in the case $N = 4$, a group containing $\Gamma_0(N)$). One must therefore look at the whole collection of related functions $\Theta_{Q,\delta}(\tau)$ obtained by replacing the sum over $\mathbb{Z}^r$ in the definition of Θ_Q by a sum over the shifted lattice $\delta + \mathbb{Z}^r$ ($\delta \in N^{-1}\mathbb{Z}^r \cap A^{-1}\mathbb{Z}^r$), apply the Poisson summation formula to each one, and then analyze in detail the transformation behaviour of the whole collection of functions under $SL_2(\mathbb{Z})$; the details are quite tedious and we will not give them.

We end this section by looking in more detail at the unimodular case for which we proved the transformation law of the theta series. As for the theta series θ^2, θ^4, θ^8 considered above, we get as an immediate consequence of the transformation law explicit formulas for the numbers of representations $r_Q(n)$. For instance, if Q is the so-called "E_8" form

$$E_8(x_1,\dots,x_8) = \sum_{i=1}^{8} x_i^2 + \sum_{i=2}^{7} x_{i-1}x_i + x_3x_8$$

(the name comes from Lie algebra theory), for which the associated even symmetric matrix is unimodular (check this!), then Θ_Q belongs to the 1-dimensional space $M_4(SL_2(\mathbb{Z})) = \langle G_4\rangle$ and starts with 1 ($r_Q(0)=1$ for any positive definite form Q), so must equal $240G_4$, and we get the highly non-obvious relation $r_{E_8}(n) = 240\sum_{d|n} d^3$ for the number of representations of a natural number n by E_8. Of course, the same would be true for any unimodular Q of rank 8, but in fact this gives no new information since it is known that all such forms are equivalent to E_8 by a change of basis in $\mathbb{Z}^8$. For $r=16$, on the other hand, it is known that there are (exactly) two inequivalent unimodular forms Q_1 and Q_2 (one of them of course being $E_8(x_1,\dots,x_8)+E_8(x_9,\dots,x_{16})$), and now the fact that $M_8(SL_2(\mathbb{Z}))$ is one-dimensional implies that these two forms represent all integers the same number of times: $r_{Q_1}(n) = r_{Q_2}(n) = 480\sum_{d|n} d^7$ ($\forall n>0$). As was observed by John Milnor, this has as the consequence the construction of a very simple counterexample in differential geometry: the two 16-dimensional tori obtained by dividing $\mathbb{R}^{16}$ by the lattices corresponding to Q_1 and Q_2 are isospectral (i.e., have the same eigenvalues of the Laplace operators; this is because these eigenvalues are determined by the lengths of closed geodesics, and each torus has $480\sigma_7(n)$ closed geodesics of length $\sqrt{n}$) but not isometric. For $r=24$ there are known to be exactly 24 non-isomorphic unimodular quadratic forms, but now the dimension of $M_{r/2}(\Gamma_1)$ is no longer 1 and they no longer all have the same theta-series. One of these forms is the one associated to the so-called **Leech lattice**, a highly symmetrical 24-dimensional lattice with many applications in group theory, coding theory and physics; it is characterized by the fact that there are no vectors x of length 1, so that the theta-series begins $1+0q+\dots$ and hence equals $(240G_4)^3 - 720\Delta$. For $r=32$ the classification of all unimodular Q has not been undertaken, but it is known that there are more than 80000000 equivalence classes. This is a consequence of a theorem of Siegel which tells us that a certain weighted sum of the theta-series attached to all the Q of a given dimension r is equal to the Eisenstein series $G_{r/2}$. Finally, we observe that the fact that Θ_Q for Q unimodular is a modular form of weight $k=r/2$ on the full modular group and has constant term 1 implies that it equals $c_k^{-1}G_k$ plus a cusp form and hence, by the estimate given in §2.3, that $r_Q(n) = c_k^{-1}\sigma_{k-1}(n) + \mathrm{O}(n^{k/2})$ as $n\to\infty$, where c_k is the constant term of the Eisenstein series G_k. This makes it obvious why r must be divisible by 8, since $r_Q(n)$ is intrinsically nonnegative but $c_k<0$ when $k\equiv 2 \pmod 4$. There also exists a purely algebraic proof that the rank of a positive definite even unimodular matrix is always divisible by 8, but it is quite subtle and—once one has gotten used to theta series—probably less natural than the one provided by the modular theory.

3.2. Theta series with spherical coefficients. One can generalize theta series by including so-called spherical functions as coefficients. If $Q:\mathbb{Z}^r\to\mathbb{Z}$ is our quadratic form, then a homogeneous polynomial $P(x)=P(x_1,\dots,x_r)$ is **spherical with respect to Q** if $\Delta_Q P=0$, where Δ_Q is the Laplace operator for Q (i.e. $\Delta_Q=\sum_j \frac{\partial^2}{\partial y_j^2}$ in a coordinate system $\{y_j\}$ for which $Q=\sum y_j^2$, or $\Delta_Q=2(\frac{\partial}{\partial x_1},\dots,\frac{\partial}{\partial x_r})A^{-1}(\frac{\partial}{\partial x_1},\dots,\frac{\partial}{\partial x_r})^t$ in the original coordinate system $\{x_j\}$, where $Q(x)=\frac{1}{2}x^tAx$). If P is such a function, say of degree ν, then the generalized theta-series

$$\Theta_{Q,P}(\tau)=\sum_{x\in\mathbb{Z}^r}P(x)\,q^{Q(x)}$$

is a modular form of weight $\frac{r}{2}+\nu$ (and of the same level and character as for $P\equiv 1$), and is a cusp form if $\nu>0$. As an example, let $Q(x_1,x_2)=x_1^2+x_2^2$, $\Delta_Q=\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}$, $P(x_1,x_2)=x_1^4-6x_1^2x_2^2+x_2^4$; then $\frac{1}{4}\Theta_{Q,P}=q-4q^2+0q^3+16q^4-14q^5+\cdots$ belongs to the one-dimensional space $S_5(\Gamma_0(4),\chi_4)$ and hence is equal to the function

$$\Delta(\tau)^{1/6}\Delta(2\tau)^{1/12}\Delta(4\tau)^{1/6}=q\prod_{n=1}^{\infty}(1-q^n)^{2+2\gcd(n,4)}.$$

That $P(x)$ here is the real part of $(x_1+ix_2)^4$ is no accident: in general, all spherical polynomials of degree ν can be obtained as linear combinations of the special spherical functions $(\zeta^tAx)^\nu$, where $\zeta\in\mathbb{C}^r$ is isotropic (i.e., $Q(\zeta)=\frac{1}{2}\zeta^tA\zeta=0$). Still more generally, one can generalize theta series by adding congruence conditions to the summation over $x\in\mathbb{Z}^r$ or, equivalently, by multiplying the spherical function $P(x)$ by some character or other periodic function of x. An example is given by the identity of Freeman Dyson mentioned in Chapter 0.

Exercises

1. Check that (1), or more generally (2), is compatible with multiplication, i.e. a function f which satisfies this relation for two elements of $\Gamma_0(4)$ (resp. $\Gamma_0(N)$) also does so for their product.

2. Show that the group $\Gamma_0(4)$ is generated by the two transformations $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1&0\\4&1\end{smallmatrix}\right)$. (Hint: show that $|a|+|c|$ for a general matrix $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\Gamma(4)$ can be made smaller by multiplying this matrix on the left by one of these two matrices or their inverses.)

3. Show that the "theta group" Γ_θ generated by $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}0&-1/2\\2&0\end{smallmatrix}\right)$ (i.e., by the two transformations $\tau\mapsto\tau+1$ and $\tau\mapsto -1/4\tau$ under which $\theta(\tau)$ has simple transformation properties) is given by

$$\Gamma_\theta=\left\{\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\begin{pmatrix}\mathbb{Z}&\mathbb{Z}\\4\mathbb{Z}&\mathbb{Z}\end{pmatrix}\cup\begin{pmatrix}2\mathbb{Z}&\frac{1}{2}\mathbb{Z}\\2\mathbb{Z}&2\mathbb{Z}\end{pmatrix}\;\middle|\;ad-bc=1\right\};$$

in particular, Γ_θ contains $\Gamma_0(4)$ as a subgroup of index 2.

4. Show that $\theta(\tau)^2$ transforms under Γ_θ by $\theta\big(\frac{a\tau+b}{c\tau+d}\big)^2=\varepsilon\,(c\tau+d)\,\theta(\tau)^2$ for $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\Gamma_\theta$, where $\varepsilon(\gamma)$ is i^{d-1} if $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\left(\begin{smallmatrix}\mathbb{Z}&\mathbb{Z}\\4\mathbb{Z}&\mathbb{Z}\end{smallmatrix}\right)$ and i^{2b} if $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\left(\begin{smallmatrix}2\mathbb{Z}&\frac{1}{2}\mathbb{Z}\\2\mathbb{Z}&2\mathbb{Z}\end{smallmatrix}\right)$.

*5. Show that if A is an even symmetric $2k \times 2k$ matrix, then $D = (-1)^k \det(A)$ is congruent to 0 or 1 modulo 4. (Hint: Show that $D \equiv \big(\sum_\tau \prod_{i<\tau(i)} a_{i\,\tau(i)}\big)^2 \pmod 4$, where the sum runs over all permutations of $\{1, 2, \ldots, 2k\}$ which are free involutions, i.e. $\tau(\tau(i)) = i$, $\tau(i) \neq i$ for all i.)

6. Let A be a positive definite symmetric $r \times r$ matrix with real coefficients. Show that the Fourier transform of $f(x) = e^{-\pi x^t A x}$ $(x \in \mathbb{R}^r)$ is $\tilde{f}(x) = \det(A)^{-\frac{1}{2}}\, e^{-\pi x^t A^{-1} x}$. (Hint: By a change of basis in $\mathbb{R}^r$ one can reduce to the case when A is diagonal.)

The Rankin-Selberg convolution method is one of the most powerful tools in the theory of automorphic forms. In this chapter we explain two principal variants of it—one involving non-holomorphic Eisenstein series and one involving only the holomorphic Eisenstein series constructed in 1.1. We will also give several applications. The essential ingredients of the Rankin-Selberg method are various types of Eisenstein series, and we begin by studying the main properties of some of these.

4.1. Non-holomorphic Eisenstein series. For $\tau = u + iv \in \mathfrak{H}$ and $s \in \mathbb{C}$ define

$$G(\tau, s) = \frac{1}{2} \sum_{m,n}{}' \frac{\Im(\tau)^s}{|m\tau + n|^{2s}}, \tag{1}$$

(sum over $m, n \in \mathbb{Z}$ not both zero). The series converges absolutely and locally uniformly for $\Re(s) > 1$ and defines a function which is Γ_1-invariant in τ for the same reason that G_k in Chapter 1 was a modular form. As a sum of pure exponential functions, it is a holomorphic function of s in the same region, but, owing to the presence of $v = \Im(\tau)$ and the absolute value signs, it is not holomorphic in τ. The function $G(\tau, s)$ is known in the literature under both the names "non-holomorphic Eisenstein series" and "Epstein zeta function" (in general, the Epstein zeta function of a positive definite quadratic form Q in r variables is the Dirichlet series $\sum'_{x \in \mathbb{Z}^r} Q(x)^{-s}$; if $r = 2$, then this equals $2d^{-s/2}G(\tau, s)$ where $-d$ is the discriminant of Q and τ the root of $Q(z, 1) = 0$ in the upper half plane). Its main properties, besides the Γ_1-invariance, are summarized in

Proposition. *The function $G(\tau, s)$ can be meromorphically extended to a function of s which is entire except for a simple pole of residue $\frac{\pi}{2}$ (independent of τ !) at $s = 1$. The function $G^*(\tau, s) = \pi^{-s}\Gamma(s)G(\tau, s)$ is holomorphic except for simple poles of residue $\frac{1}{2}$ and $-\frac{1}{2}$ at $s = 1$ and $s = 0$, respectively, and satisfies the functional equation $G^*(\tau, s) = G^*(\tau, 1 - s)$.*

Proof. We sketch two proofs of this. The first is analogous to Riemann's proof of the functional equation of $\zeta(s)$. For $\tau = u + iv \in \mathfrak{H}$ let Q_τ be the positive definite binary quadratic form $Q_\tau(m, n) = v^{-1}|m\tau + n|^2$ of discriminant -4 and $\Theta_\tau(t) = \sum_{m,n \in \mathbb{Z}} e^{-\pi Q_\tau(m,n)t}$ the associated theta series. The Mellin transformation formula (cf. Appendix **A2**) implies

$$G^*(\tau, s) = \frac{1}{2}\Gamma(s) \sum_{m,n}{}' \bigl[\pi Q_\tau(m, n)\bigr]^{-s} = \frac{1}{2} \int_0^\infty \bigl(\Theta_\tau(t) - 1\bigr) t^{s-1}\, dt.$$

On the other hand, the Poisson summation formula (cf. Appendix **A1**) implies that $\Theta_\tau(\frac{1}{t}) = t\Theta_\tau(t)$, so the function $\phi(t) = \frac{1}{2}\bigl(\Theta_\tau(t) - 1\bigr)$ satisfies $\phi(t^{-1}) = -\frac{1}{2} + \frac{1}{2}t + t\phi(\frac{1}{t})$. The "functional equation principle" formulated in Appendix **A2** now gives the assertions of the theorem.

The second proof, which requires more calculation, but also gives more information, is to compute the Fourier development of $G(\tau,s)$. The computation is very similar to that for G_k in Chapter 1, so we can be brief. Splitting up the sum defining $G(\tau,s)$ into the terms with $m=0$ and those with $m\neq 0$, and combining each summand with its negative, we find

$$G(\tau,s)=\zeta(2s)v^s+v^s\sum_{m=1}^{\infty}\biggl(\sum_{n=-\infty}^{\infty}|m\tau+n|^{-2s}\biggr)\qquad\qquad(\tau=u+iv).$$

Substituting into this formula (3) of Appendix **A1**, we find

$$\begin{aligned}G(\tau,s)=\zeta(2s)v^s+&\ \frac{\pi^{\frac12}\Gamma(s-\frac12)}{\Gamma(s)}v^{1-s}\sum_{m=1}^{\infty}m^{1-2s}\\&+\frac{2\pi^s}{\Gamma(s)}v^{\frac12}\sum_{\substack{m\geq1\\ r\neq0}}m^{\frac12-s}|r|^{s-\frac12}\,K_{s-\frac12}(2\pi m|r|v)\,e^{2\pi imru},\end{aligned}$$

where $K_\nu(t)$ is the K-Bessel function $\int_0^\infty e^{-t\cosh u}\cosh(\nu u)\,du$. Hence

$$G^*(\tau,s)=\zeta^*(2s)v^s+\zeta^*(2s-1)v^{1-s}+2v^{\frac12}\sum_{n\neq0}\sigma^*_{s-\frac12}(|n|)K_{s-\frac12}(2\pi|n|v)e^{2\pi inu},$$

where $\zeta^*(s)$ denotes the meromorphic function $\pi^{-s/2}\Gamma(s/2)\zeta(s)$ and $\sigma^*_\nu(n)$ the arithmetic function $|n|^\nu\sum_{d|n}d^{-2\nu}$. The analytic continuation properties of G^* now follow from the facts that $\zeta^*(s)$ is holomorphic except for simple poles of residue 1 and -1 at $s=1$ and $s=0$, respectively, that $\sigma^*_\nu(n)$ is an entire function of ν, and that $K_\nu(t)$ is entire in ν and exponentially small in t as $t\to\infty$, while the functional equation follows from the functional equations $\zeta^*(1-s)=\zeta^*(s)$ (cf. Appendix **A2**), $\sigma^*_{-\nu}(n)=\sigma^*_\nu(n)$, and $K_{-\nu}(t)=K_\nu(t)$.

As an immediate consequence of the Fourier development of G^* and the identity $K_{\frac12}(t)=\sqrt{\pi/2t}\,e^{-t}$, we find

$$\begin{aligned}\lim_{s\to1}\biggl(G^*(\tau,s)-\frac{1/2}{s-1}\biggr)&=\frac{\pi}{6}v-\frac12\log v+C+2\sum_{m,r=1}^{\infty}\frac1m\,\Re(e^{2\pi imr\tau})\\&=\frac{\pi}{6}v-\frac12\log v+C-\sum_{r=1}^{\infty}\log|1-e^{2\pi ir\tau}|^2\\&=-\frac{1}{24}\log\bigl(v^{12}|\Delta(\tau)|^2\bigr)\;+\;C,\end{aligned}$$

where $C=\lim_{s\to1}(\zeta^*(s)-(s-1)^{-1})$ is a certain constant (in fact given by $\frac12\gamma-\frac12\log4\pi$, where γ is Euler's constant) and $\Delta(\tau)$ the discriminant function of 1.2. This formula is called the **Kronecker limit formula** and has many applications in number theory. Together with the invariance of $G(\tau,s)$ under $SL_2(\mathbb{Z})$, it leads to another proof of the modular transformation property of $\Delta(\tau)$.

4.2. The Rankin-Selberg method (non-holomorphic case) and applications. In this section we describe the "unfolding method" invented by Rankin and Selberg in their papers of 1939–40. Suppose that $F(\tau)$ is a smooth Γ_1-invariant function in the upper half-plane and tends to 0 rapidly (say, exponentially) as $v = \Im(\tau) \to \infty$. (In the original papers of Rankin and Selberg, $F(\tau)$ was the function $v^{12}|\Delta(\tau)|^2$.) The Γ_1-invariance of F implies in particular the periodicity property $F(\tau+1) = F(\tau)$ and hence the existence of a Fourier development $F(u+iv) = \sum_{n\in\mathbb{Z}} c_n(v)e^{2\pi inu}$. We define the **Rankin-Selberg transform** of F as the Mellin transform (cf. Appendix **A2**) of the constant term $c_0(v)$ of F:

$$R(F;s) = \int_0^\infty c_0(v)\, v^{s-2}\, dv \tag{1}$$

(notice that there is a shift of s by 1 with respect to the usual definition of the Mellin transform). Since $F(u+iv)$ is bounded for all v and very small as $v \to \infty$, its constant term

$$c_0(v) = \int_0^1 F(u+iv)\, du \tag{2}$$

also has these properties. Hence the integral in (1) converges absolutely for $\Re(s) > 1$ and defines a holomorphic function of s in that domain.

Theorem. *The function $R(F;s)$ can be meromorphically extended to a function of s and is holomorphic in the half-plane $\Re(s) > \frac{1}{2}$ except for a simple pole of residue $\kappa = \frac{3}{\pi}\iint_{\mathfrak{H}/\Gamma_1} F(\tau)\, d\mu$ at $s = 1$. The function $R^*(F;s) = \pi^{-s}\Gamma(s)\zeta(2s)R(F;s)$ is holomorphic everywhere except for simple poles of residue $\pm\frac{\pi}{6}\kappa$ at $s = 1$ and $s = 0$ and $R^*(F;s) = R^*(F;1-s)$.*

(Recall that $d\mu$ denotes the $SL(2,\mathbb{R})$-invariant volume measure $v^{-2}\, du\, dv$ on $\mathfrak{H}/\Gamma_1$ and that the area of $\mathfrak{H}/\Gamma_1$ with respect to this measure is $\pi/3$; thus κ is simply the average value of F in the upper half-plane.)

Proof. We will show that $\zeta(2s)R(F;s)$ is equal to the Petersson scalar product of $\overline{F}$ with the non-holomorphic Eisenstein series of the last section:

$$\zeta(2s)R(F;s) = \iint_{\mathfrak{H}/\Gamma_1} G(\tau,s)\, F(\tau)\, d\mu. \tag{3}$$

The assertions of the theorem then follow immediately from the proposition in that section.

To prove (3) we use the method called "unfolding" (sometimes also referred to as the "Rankin-Selberg trick"). Let Γ_∞ denote the subgroup $\left\{\pm\left(\begin{smallmatrix}1 & n\\ 0 & 1\end{smallmatrix}\right) \mid n\in\mathbb{Z}\right\}$ of Γ_1 (the "∞" in the notation refers to the fact that Γ_∞ is the stabilizer in Γ_1 of infinity). The left cosets of Γ_∞ in Γ_1 are in 1:1 correspondence with pairs of coprime integers (c,d), considered up to sign: multiplying a matrix $\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)$ on the left by $\left(\begin{smallmatrix}1 & n\\ 0 & 1\end{smallmatrix}\right)$ produces a new matrix with the same second row, and any two matrices with the same second row are related in this way. Also,

$\Im(\gamma(\tau)) = v/|c\tau+d|^2$ for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1$. Finally, any non-zero pair of integers (m,n) can be written uniquely as (rc, rd) for some $r > 0$ and coprime c and d. Hence for $\Re(s) > 1$ we have

$$G(\tau,s) = \frac{1}{2}\sum_{r=1}^{\infty} \sum_{c,d \text{ coprime}} \frac{\Im(\tau)^s}{|r(c\tau+d)|^{2s}} = \zeta(2s) \sum_{\gamma\in\Gamma_\infty\backslash\Gamma_1} \Im\bigl(\gamma(\tau)\bigr)^s.$$

Therefore, denoting by $\mathcal{F}$ a fundamental domain for the action of Γ_1 on $\mathfrak{H}$, and observing that the sum and integral are absolutely convergent and that both F and $d\mu$ are Γ_1-invariant, we obtain

$$\begin{aligned} \zeta(2s)^{-1} \iint\limits_{\mathfrak{H}/\Gamma_1} G(\tau,s)\, F(\tau)\, d\mu &= \iint\limits_{\mathcal{F}} \sum_{\gamma\in\Gamma_\infty\backslash\Gamma_1} \Im(\gamma\tau)^s\, F(\gamma\tau)\, d\mu \\ &= \sum_{\gamma\in\Gamma_\infty\backslash\Gamma_1} \iint\limits_{\gamma\mathcal{F}} \Im(\tau)^s\, F(\tau)\, d\mu. \end{aligned}$$

Notice that we have spoiled the invariance of the original representation: both the fundamental domain and the set of coset representatives for $\Gamma_\infty\backslash\Gamma_1$ must be chosen explicitly for the individual terms in what we have just written to make sense. Now comes the unfolding argument: the different translates $\gamma\mathcal{F}$ of the original fundamental domain are disjoint, and they fit together exactly to form a fundamental domain for the action of Γ_∞ on $\mathfrak{H}$ (here we ignore questions about the boundaries of the fundamental domains, since these form a set of measure zero and can be ignored.) Hence finally

$$\zeta(2s)^{-1} \iint_{\mathfrak{H}/\Gamma_1} G(\tau,s)\, F(\tau)\, d\mu = \iint_{\mathfrak{H}/\Gamma_\infty} \Im(\tau)^s\, F(\tau)\, d\mu.$$

Since the action of Γ_∞ on $\mathfrak{H}$ is given by $u \mapsto u+1$, the right-hand side of this can be rewritten as $\int_0^\infty \bigl(\int_0^1 F(u+iv)\,du\bigr)\, v^{s-2}\,dv$, and in view of equation (2) this is equivalent to the assertion (3). A particularly pleasing aspect of the computation is that—unlike the usual situation in mathematics where a simplification at one level of a formula must be paid for by an increased complexity somewhere else—the unfolding simultaneously permitted us to replace the complicated infinite sum defining the Eisenstein series by a single term $\Im(\tau)^s$ *and* to replace the complicated domain of integration $\mathfrak{H}/\Gamma_1$ by the much simpler $\mathfrak{H}/\Gamma_\infty$ and eventually just by $(0,\infty)$.

We now give some applications of the theorem. The first application is to the Γ_1-invariant function $F(\tau) = v^k|f(\tau)|^2$, where $f = \sum a(n)q^n$ is any cusp form in S_k (in the original papers of Rankin and Selberg, as already mentioned, f was the discriminant function of §1.2, $k=12$). We have

$$F(u+iv) = v^k \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a(n)\,\overline{a(m)}\, e^{2\pi i(n-m)u}\, e^{-2\pi(n+m)v}$$

and hence $c_0(v) = v^k \sum_{n=1}^{\infty} |a(n)|^2 \, e^{-4\pi n v}$. Therefore

$$R(F;s) = \sum_{n=1}^{\infty} |a(n)|^2 \int_0^{\infty} v^k \, e^{-4\pi n v} \, v^{s-2} \, dv = \frac{\Gamma(s+k-1)}{(4\pi)^{s+k-1}} \sum_{n=1}^{\infty} \frac{|a(n)|^2}{n^{s+k-1}}.$$

This proves the meromorphic continuability and functional equation of the "Rankin zeta function" $\sum |a(n)|^2 n^{-s}$; moreover, applying the statement about residues in the theorem and observing that κ here is just $3/\pi$ times the Petersson scalar product of f with itself, we find

$$(f,f) = \frac{\pi}{3} \frac{(k-1)!}{(4\pi)^k} \operatorname{Res}_{s=1}\biggl(\sum_{n=1}^{\infty} \frac{|a(n)|^2}{n^{s+k-1}}\biggr). \tag{4}$$

If f is a Hecke form, then the coefficients $a(n)$ real and $\sum \dfrac{a(n)^2}{n^{s+k-1}} = \zeta(s) \sum \dfrac{a(n^2)}{n^{s+k-1}}$ by an easy computation using the shape of the Euler product of the L-series of f, so this can be rewritten in the equivalent form

$$(f,f) = \frac{\pi}{3} \frac{(k-1)!}{(4\pi)^k} \sum_{n=1}^{\infty} \frac{a(n^2)}{n^s} \bigg|_{s=k}. \tag{5}$$

As a second application, we get a proof different from the usual one of the fact that the Riemann zeta function has no zeros on the line $\Re(s) = 1$; this fact is one of the key steps in the classical proof of the prime number theorem. Indeed, suppose that $\zeta(1+i\alpha) = 0$ for some real number α (necessarily different from 0), and let $F(\tau)$ be the function $G\bigl(\tau, \frac{1}{2}(1+i\alpha)\bigr)$. Since both $\zeta(2s)$ and $\zeta(2s-1)$ vanish at $s = \frac{1}{2}(1+i\alpha)$ (use the functional equation of ζ!), the formula for the Fourier expansion of $G(\tau,s)$ proved in the last section shows that $F(\tau)$ is exponentially small as $v \to \infty$ and has a constant term $c_0(v)$ which vanishes identically. Therefore the Rankin-Selberg transform $R(F;s)$ is zero for $\Re(s)$ large, and then by analytic continuation for all s. But we saw above that $R(F;s)$ is the integral of $F(\tau)$ against $G(\tau,s)$, so taking $s = \frac{1}{2}(1-i\alpha)$, $G(\tau,s) = \overline{F(\tau)}$, we find that the integral of $|F(\tau)|^2$ over $\mathfrak{H}/\Gamma_1$ is zero. This is impossible since $F(\tau)$ is clearly not identically zero.

Finally, we can re-interpret the statement of the Rankin-Selberg identity in more picturesque ways. Suppose that we knew that the constant term $c_0(v)$ of F had an asymptotic expansion $c_0(v) = C_0 v^{\lambda_0} + C_1 v^{\lambda_1} + C_2 v^{\lambda_2} + \cdots$ as v tends to 0. Then breaking up the integral in the definition of $R(F;s)$ into the part from 0 to 1 and the part from 1 to infinity, and observing that the second integral is convergent for all s, we would discover that $R(F;s)$ has simple poles of residue C_j at $s = 1-\lambda_j$ for each j and no other poles. Similarly, a term $Cv^{\lambda}(\log v)^{m-1}$ would correspond to an mth order pole of $R(F;s)$ at $1-\lambda$. But the theorem tells us that $R(F;s)$ has a simple pole of residue κ at $s = 1$ and otherwise poles only at the values $s = \frac{1}{2}\rho$, where ρ is a non-trivial zero of the Riemann zeta function. It is thus reasonable to think, and presumably under suitable hypotheses possible to prove, that $c_0(v)$ has an asymptotic expansion as $v \to 0$ consisting of one constant term κ and a sum of terms $C_\rho v^{1-\rho/2}$ for the various zeros of $\zeta(s)$. Assuming the Riemann hypothesis, these latter terms are of the form $v^{3/4}$ times an oscillatory function $A\cos(\frac{1}{2}\Im(\rho)\log v + \phi)$ for some amplitude A and phase ϕ. Figure 1 illustrates this behavior for the constant term $v^{12}\sum \tau(n)^2 e^{-4\pi n v}$ of $v^{12}|\Delta(\tau)|^2$; the predicted oscillatory behavior is clearly visible, and a rough measurement of the period of the primary oscillation leads to a rather accurate estimate of the imaginary part of the smallest non-trivial zero of $\zeta(s)$.

Fig. 1. The constant term $c_0(v) = v^{12}\sum_{n=1}^{\infty}\tau(n)^2\, e^{-4\pi n v}$

In a related vein, we see that the difference between $c_0(v)$ and the average value κ of F for small v should be estimated by $\mathrm{O}\big(v^{\frac{1}{4}+\epsilon}\big)$ if the Riemann hypothesis is true and by $\mathrm{O}\big(v^{\frac{1}{2}+\epsilon}\big)$ unconditionally. Since $c_0(v)$ is simply the average value of $F(\tau)$ along the unique closed horocycle of length v^{-1} in the Riemannian manifold $\mathfrak{H}/\Gamma_1$, and since F is an essentially arbitrary function on this manifold, we can interpret this as a statement about the uniformity with which the closed horocycles on $\mathfrak{H}/\Gamma_1$ fill it up as their length tends to infinity.

4.3. The Rankin-Selberg method (holomorphic case). The calculations here are very similar to those of Section 4.2, so we can be fairly brief. Let $f(\tau) = \sum_{n=1}^{\infty} a(n)q^n$ be a cusp form of weight k on Γ and $g(\tau) = \sum_{n=0}^{\infty} b(n)q^n$ a modular form of some smaller weight l. We assume for the moment that $k - l > 2$, so that there is a holomorphic Eisenstein series G_{k-l} of weight $k-l$. Our object is to calculate the scalar product of $f(\tau)$ with the product $G_{k-l}(\tau)g(\tau)$.

Ignoring convergence problems for the moment, we find (with $h = k - l$)

$$G_h(\tau) = \frac{(h-1)!}{(2\pi i)^h}\,\frac{1}{2}\sum_{m,n}{}' \frac{1}{(m\tau+n)^h} = \frac{(h-1)!}{(2\pi i)^h}\,\zeta(h) \sum_{\left(\begin{smallmatrix} \cdot & \cdot \\ c & d \end{smallmatrix}\right)\in\Gamma_\infty\backslash\Gamma_1} \frac{1}{(c\tau+d)^h},$$

whence

$$\frac{(2\pi i)^h\, v^k}{(h-1)!\zeta(h)}\, f(\tau)\,\overline{G_h(\tau)\,g(\tau)} = \sum_{\left(\begin{smallmatrix} \cdot & \cdot \\ c & d \end{smallmatrix}\right)\in\Gamma_\infty\backslash\Gamma_1} \frac{v^k}{|c\tau+d|^{2k}}\,(c\tau+d)^k f(\tau)\,\overline{(c\tau+d)^l g(\tau)}$$
$$= \sum_{\gamma\in\Gamma_\infty\backslash\Gamma_1} \Im(\gamma\tau)^k f(\gamma\tau)\overline{g(\gamma\tau)},$$

and consequently, by the same unfolding argument as in 4.2,

$$\frac{(2\pi i)^h}{(h-1)!\zeta(h)}\,(f, G_h\cdot g) = \iint_{\mathcal{F}} \sum_{\gamma\in\Gamma_\infty\backslash\Gamma_1} \Im(\gamma\tau)^k\, f(\gamma\tau)\,\overline{g(\gamma\tau)}\,d\mu$$
$$= \sum_{\gamma\in\Gamma_\infty\backslash\Gamma_1} \iint_{\gamma\mathcal{F}} \Im(\tau)^k\, f(\tau)\,\overline{g(\tau)}\,d\mu$$
$$= \int_0^\infty \Bigl(\int_0^1 f(u+iv)\,\overline{g(u+iv)}\,du\Bigr)\, v^{k-2}\,dv$$
$$= \int_0^\infty \Bigl(\sum_{n=1}^\infty a(n)\,\overline{b(n)}\,e^{-4\pi n v}\Bigr)\, v^{k-2}\,dv$$
$$(1)\qquad = \frac{(k-2)!}{(4\pi)^{k-1}} \sum_{n=1}^\infty \frac{a(n)\,\overline{b(n)}}{n^{k-1}}.$$

In other words, the scalar product of f and $G_h\cdot g$ is up to a simple factor equal to the value at $s=k-1$ of the convolution of the L-series of f and $\overline{g}$. The various steps in the calculation will be justified if $\iint_{\Gamma_\infty\backslash\mathfrak{H}} |f(\tau)g(\tau)|v^k\,d\mu$ converges. Since $f(\tau) = \mathrm{O}(v^{-k/2})$ and $g(\tau)=\mathrm{O}(v^{-l})$, this will certainly be the case if $k>2l+2$.

We can generalize the computation just done by replacing the product $G_h\cdot g$ by the function $F_\nu(G_h,g)$ defined in Section 1.3, where now $h+l+2\nu=k$. Here we find

$$\frac{(2\pi i)^h}{(h-1)!\zeta(h)}\,F_\nu(G_h,g) = F_\nu\Bigl(\sum_{\gamma\in\Gamma_\infty\backslash\Gamma_1}\frac{1}{(c\tau+d)^h},\, g(\tau)\Bigr)$$
$$= \bigl(\frac{i}{2\pi}\bigr)^\nu \sum_\gamma \sum_{\mu=0}^{\nu} \frac{(h+\nu-1)!(l+\nu-1)!}{\mu!(\nu-\mu)!(h-1)!(l+\mu-1)!}\,\frac{c^{\nu-\mu}g^{(\mu)}(\tau)}{(c\tau+d)^{h+\nu-\mu}}$$
$$= \bigl(\frac{i}{2\pi}\bigr)^\nu \binom{h+\nu-1}{\nu} \sum_\gamma \frac{g^{(\nu)}(\gamma\tau)}{(c\tau+d)^k},$$

where in the last line we have used formula (11) of 1.3. The same argument as before now leads to

$$\frac{(2\pi i)^h}{(h-1)!\zeta(h)}\,(f, F_\nu(G_h,g)) = \frac{\binom{h+\nu-1}{\nu}}{(-2\pi i)^\nu} \int_0^\infty \int_0^1 f(\tau)\,\overline{g^{(\nu)}(\tau)}\,du\,v^{k-2}\,dv$$
$$(2)\qquad = \binom{h+\nu-1}{\nu}\frac{(k-2)!}{(4\pi)^{k-1}} \sum_{n=1}^\infty \frac{a(n)\,\overline{b(n)}}{n^{k-\nu-1}},$$

the steps being justified this time if $k > 2l+2\nu+2$ or $h > l+2$. Again the result is that the Petersson scalar product in question is proportional to a special value of the convolution of the L-series of f and g.

We now specialize to the case that $g = G_l$ is the (normalized) Eisenstein series of weight l and f a (normalized) Hecke eigenform. Then $\overline{b(n)} = b(n) = \sigma_{l-1}(n)$. An easy calculation (Exercise 1) shows that

$$\sum_{n=1}^{\infty} \frac{a(n)\,\sigma_{l-1}(n)}{n^s} = \frac{L(f,s)\,L(f,s-l+1)}{\zeta(2s-l-k+2)} \qquad \Bigl(\Re(s) > l + \frac{k-1}{2}\Bigr)\,, \tag{3}$$

where $L(f,s) = \sum_{n=1}^{\infty} a(n)n^{-s}$ is the Hecke L-series of f. Specializing this to $s = k-\nu-1$ and replacing the L-series $L(f,s)$ by $L^*(f,s)$ to absorb superfluous gamma factors and powers of π, we obtain

$$\bigl(f,\, F_\nu(G_h, G_l)\bigr) = \frac{(-1)^{h/2}}{2^{k-1}} \binom{k-2}{\nu} L^*(f, k-\nu-1)\, L^*(f, h+\nu)\,, \tag{4}$$

in which the right-hand side is symmetric in h and l (as of course it must be) by virtue of the functional equation of $L^*(f,s)$. We will apply this identity in the next chapter to deduce algebraicity results about the special values of the L-series of f at integer arguments between 0 and k.

Exercises

1. Fill in the details of the two proofs of the functional equation of $G(\tau,s)$ sketched in 4.1, using the information in Appendices **A1** and **A2**.

2. Let $f(\tau) = \sum_{n=1}^{\infty} a(n)q^n$ be a Hecke form of weight k. Prove the identity $\sum a(n)^2 n^{-s} = \zeta(s-k+1) \sum a(n^2)n^{-s}$ (cf. 4.1, passage from (4) to (5)) and the identity (4) of 4.3. (Hint: Prove the corresponding identitites for each factor of the Euler products of the Dirichlet series in question.)

3. For n a nonnegative integer and $k > 2$ define **the nth Poincaré series of weight k** by $P_n(\tau) = \sum_\gamma (c\tau+d)^{-k}\, e^{2\pi i \gamma(\tau)}$, where $\gamma = \bigl(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\bigr)$ runs over (representatives) of classes of $\Gamma_\infty \backslash \Gamma_1$ and $\gamma(\tau) = (a\tau+b)/(c\tau+d)$ as usual. (For $n = 0$ this reduces to a multiple of the usual Eisenstein series.) Show that the series converges absolutely and defines a modular form of weight k on Γ_1 whose scalar product with an arbitrary cusp form $f(\tau) = \sum_{n=1}^{\infty} a(n)q^n \in S_k(\Gamma_1)$ equals $(k-2)!a(n)/(4\pi n)^{k-1}$.

4. Let $P_n(\tau)$ be the Poincaré series defined in the preceding exercise, and let $g(\tau) = \sum_{n=1}^{\infty} b(n)q^n \in S_l$ and $k = h + l + 2\nu$ as in §4.3. Show that the function $F_\nu(G_h, g)$ is a multiple of $\sum_{n=0}^{\infty} n^\nu b(n) P_n(\tau)$.

If $f(\tau)$ is a cusp form of weight k on $\Gamma = SL_2(\mathbb{Z})$, we define the **periods** of f as the $k-1$ numbers

$$(1) \qquad r_n(f) = \int_0^{i\infty} f(\tau)\,\tau^n\,d\tau \qquad (n = 0,\,1,\,\ldots\,,k-2)\;,$$

where the integral is taken along the imaginary axis. Writing it $(t>0)$ for τ and using the integral representation for the Hecke L-series of f as given in Chapter 2, we see that this definition is equivalent to the formula

$$(2) \qquad r_n(f) = i^{n+1}L^*(f,n+1) \qquad (\,L*(f,s) = (2\pi)^{-s}\Gamma(s)L(f,s)\,)\;.$$

Now there is a conjectural general principle saying that if $L(s)$ is any Dirichlet series which arises naturally ("motivically"), has an Euler product, and satisfies a functional equation of the form $\gamma(s)L(s) = \pm\gamma(h-s)L(h-s)$, where $\gamma(s)$ is a product of gamma functions and h a positive integer, then $L(s)$ should have "special" values at all integers s_0 for which neither $\gamma(s)$ nor $\gamma(h-s_0)$ has a pole. Here "special" means that the value $L(s_0)$, when divided by an appropriate "period" (=complex number obtained as the integral over some closed cycle of some differential form with algebraic coefficients), should be an algebraic number. For instance, the Riemann zeta function has a functional equation of the stated type with $\gamma(s) = \pi^{s/2}\Gamma(s/2)$, which has poles at $s = 0,\,-2,\,-4,\,\ldots$, and $h = 1$, so we should have good special values at $s_0 = 2,\,4,\,6,\,\ldots$ and $s_0 = -1,\,-3,\,\ldots\,$, and indeed we have Euler's formulae $\zeta(n) = -\frac{(2\pi i)^n B_n}{2n!}$, $\zeta(1-n) = -\frac{B_n}{n}$ for positive even integers n.

In the case where $L(s) = L(f,s)$ is the Hecke L-series of a Hecke eigenform $f \in S_k$, the gamma factor $\gamma(s) = (2\pi)^{-s}\Gamma(s)$ has poles at $s = 0,\,-1,\,-2,\,\ldots$ and $h = k$, so we expect algebraicity results for $L(s_0)$ for $s_0 \in \{1,2,\ldots\,,k-1\}$, i.e., for the periods as defined above. This turns out to be true. In this chapter we will study these periods and see that they give an approach to modular forms that contains just as interesting information as that which one gets from the usual point of view of Fourier coefficients. In particular, it turns out that one can describe completely the relations among the periods of given parity (n even or n odd) of modular forms and that one can give an elementary and calculable description of the action of Hecke operators in terms of periods.

5.1. Period polynomials and the Eichler-Shimura isomorphism. We will always use k to denote a positive even integer and M_k (resp. S_k) for the spaces of modular (resp. cusp) forms of weight k on Γ. For $n \in 2\mathbb{Z}$, we denote by $|_n$ the action of Γ on functions given by $(f|_n\gamma)(x) = (cx+d)^{-n}f(\gamma x)$ for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, where $\gamma x = (ax+b)/(cx+d)$. For example, $f|_k\gamma = f$ for all $\gamma \in \Gamma$ if $f \in M_k$. We denote by V_k the space of polynomials of degreee $\le k-2$ in one variable, together with the action of Γ by $|_{2-k}$. If $f \in S_k$, we combine the periods $r_n(f)$ of f as defined in (1) to an element $r(f) \in V_k$ by setting

$$(3) \qquad r(f)(X) = \sum_{n=0}^{k-2} (-1)^n \binom{k-2}{n} r_n(f)\, X^{k-2-n} = \int_0^\infty f(\tau)\,(\tau - X)^{k-2}\, d\tau.$$

If we make the substitution $\tau \mapsto \gamma\tau$ in the integral, where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$, and use the identities $f|_k\gamma = f$ and $(\gamma X - \gamma\tau) = (X-\tau)/(cX+d)(c\tau+d)$, then we find that $(r(f)|_{2-k}\gamma)(X)$ is given by an integral with the same integrand as $r(f)$, but with limits $\gamma^{-1}0$ and $\gamma^{-1}\infty$ instead of 0 and ∞. In particular, for the special elements $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $U = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ of order 2 and 3 in Γ we get $r(f)|S = \int_\infty^0$ and $r(f)|U = \int_1^0$, $r(f)|U^2 = \int_\infty^1$, so $r(f) + r(f)|S$ and $r(f) + r(f)|U + r(f)|U^2$ both vanish. Thus $r(f)$ belongs to the space

$$(4) \qquad W_k = \bigl\{\phi \in V_k \;\; : \;\; \phi + \phi|S = \phi + \phi|U + \phi|U^2 = 0\bigr\}$$

(from now on we omit the subscript "$2-k$" on $|$ whenever writing the operation of 2×2 matrices on elements of V_k).

To proceed further we must look at the space of polynomials V_k in more detail. The first remark is that V_k possesses a non-degenerate invariant scalar product: if we set

$$(5) \qquad (X^n,\, X^m) = \begin{cases} (-1)^n \binom{k-2}{n}^{-1} & \text{if } n+m = k-2 \\ 0 & \text{otherwise} \end{cases}$$

for n and m between 0 and $k-2$ and extend to all of V_k by linearity, then

$$(6) \qquad (\phi|g,\, \psi|g) = (\phi,\, \psi) \qquad \text{for all} \qquad \phi,\, \psi \in V_k\,, \quad g \in PSL(2,\mathbb{R})$$

(Exercise 1). On the other hand, since the endomorphisms $\phi \mapsto \phi|S$ and $\phi \mapsto \phi|U$ of V_k have order 2 and 3, respectively, we have splittings

$$(7) \qquad V_k = A_k \oplus B_k = C_k \oplus D_k$$

for all k, where

$$A_k = \mathrm{Ker}(1-S) = \mathrm{Im}(1+S), \qquad B_k = \mathrm{Ker}(1+S) = \mathrm{Im}(1-S)$$

denote the $(+1)$- and (-1)-eigenspaces of S on V_k and

$$C_k = \mathrm{Ker}(1-U) = \mathrm{Im}(1+U+U^2), \qquad D_k = \mathrm{Ker}(1+U+U^2) = \mathrm{Im}(2-U-U^2)$$

the $(+1)$-eigenspace and the sum of the ω- and ω^2-eigenspaces of U, respectively ($\omega = e^{2\pi i/3}$). Here the operation of Γ on V_k has been extended by linearity to an action of the **group ring** $\mathbb{Z}[\Gamma]$, consisting of all finite integral linear combinations of elements of Γ, so that, for instance, $1-S$ denotes the map $\phi \mapsto \phi - \phi|S$ from V_k to V_k. Because of the invariance property (6), both splittings in (7) are orthogonal. It follows that $W_k = B_k \cap D_k$ equals $(A_k + C_k)^\perp$, the orthogonal complement in V_k of the subspace spanned by A_k and C_k. But $A_k \cap C_k = \{0\}$ for $k > 2$, since a function belonging to this intersection would be invariant under S and U and hence under the whole group Γ, a contradiction (a polynomial invariant under $US = \left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ is periodic and hence constant, and constants are not invariant under S). Hence the sum $A_k + C_k$ is in fact a direct sum and we have a direct sum decomposition

$$V_k = A_k \oplus C_k \oplus W_k\,. \tag{8}$$

In particular, $\dim(W_k)$ equals $\dim V_k - \dim A_k - \dim C_k$. But

$$\dim A_k = 1 + 2\Big[\frac{k-2}{4}\Big]\,, \qquad \dim C_k = 1 + 2\Big[\frac{k-2}{6}\Big] \tag{9}$$

(Exercise 2), so

$$\dim W_k = k - 3 - 2\Big[\frac{k-2}{4}\Big] - 2\Big[\frac{k-2}{6}\Big] = 2\dim S_k + 1\,. \tag{10}$$

Formula (10) suggests that the period mapping $r : S_k \to W_k$ is injective and even that there are two copies of S_k sitting inside W_k in a natural way. This is indeed the case. Denote by σ the matrix $\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$ and also its operation on V_k ($X^n \mapsto X^{k-2-n}$). Under the action of σ we can split V_k into the orthogonal sum $V_k^+ \oplus V_k^-$ of the $(+1)$- and (-1)-eigenspaces of σ (symmetric and anti-symmetric polynomials). Although σ has determinant -1 and hence does not lie in Γ, conjugation by σ is an automorphism of Γ. This automorphism preserves S and interchanges U and U^2, so the five subspaces A_k, B_k, C_k, D_k and W_k of V_k are all σ-invariant and hence split up into the sum of $(+1)$- and (-1)-eigenspaces. In particular, the decomposition (8) above can be refined to

$$V_k = A_k^+ \oplus A_k^- \oplus C_k^+ \oplus C_k^- \oplus W_k^+ \oplus W_k^-\,.$$

For example, for $k = 6$ this decomposition is

$$\begin{aligned}
&A_6^+ = \langle X^4 + 1, \quad X^2\rangle, \quad A_6^- = \langle X^3 - X\rangle\,,\\
&C_6^+ = \langle X^4 - 2X^3 + 3X^2 - 2X + 1\rangle, \quad C_6^- = \{0\}\,,\\
&W_6^+ = \{0\}\,, \quad W_6^- = \langle X^4 - 1\rangle\,,
\end{aligned}$$

while for $k = 12$ it is

$$\begin{aligned}
A_{12}^+ &= \langle X^{10} + 1, \quad X^8 + X^2, \quad X^6 + X^4\rangle\,,\\
A_{12}^- &= \langle X^9 - X, \quad X^7 - X^3\rangle\,,\\
C_{12}^+ &= \langle X^{10} - 5X^9 + 10X^8 - 10X^7 + 5X^6 - X^5 + 5X^4 - 10X^3 + 10X^2 - 5X + 1,\\
&\qquad\qquad X^8 - 4X^7 + 8X^6 - 10X^5 + 8X^4 - 4X^3 + X^2\rangle\,,\\
C_{12}^- &= \langle 2X^9 - 9X^8 + 16X^7 - 14X^6 + 14X^4 - 16X^3 + 9X^2 - 2X\rangle\,,\\
W_{12}^+ &= \langle 4X^9 - 25X^7 + 42X^5 - 25X^3 + 4X\rangle\,,\\
W_{12}^- &= \langle X^{10} - 1, \quad X^8 - 3X^6 + 3X^4 - X^2\rangle\,.
\end{aligned}$$

On B_k, and therefore also on the subpace W_k, the symmetric and anti-symmetric polynomials are just the odd and even polynomials, respectively, because σ is the product of S and $\left(\begin{smallmatrix}-1&0\\0&1\end{smallmatrix}\right)$ and S acts as -1. Therefore the decomposition of an element $\phi(X)\in W_k$ into its W_k^+ and W_k^- components is just the usual decomposition $\phi(X)=\frac{1}{2}(\phi(X)-\phi(-X))+\frac{1}{2}(\phi(X)+\phi(-X))$ of a polynomial into the sum of an odd and an even polynomial. Applying this to the period polynomial $\phi(X)=r(f)(X)$ of a cusp form $f\in S_k$ gives two polynomials

$$r^{\pm}(f)(X)=\pm\sum_{(-1)^n=\mp 1}\binom{k-2}{n}r_n(f)\,X^{k-2-n}\in W_k^{\pm}\,.$$

The basic result is then the following

Theorem (Eichler-Shimura isomorphism). *The map $r^+:S_k\to W_k^+$ is an isomorphism. The map $r^-:S_k\to W_k^-$ is an isomorphism onto a subspace of codimension 1, the whole space W_k^- being the sum of this subspace and the 1-dimensional space spanned by the polynomial $X^{k-2}-1$.*

Since the dimension formulas $\dim W_k^+=\dim S_k$, $\dim W_k^-=\dim S_k+1$ can be proved by a refinement of the argument proving equation (10), the essential statement to be proved is the injectivity. For this, it is enough to give a formula for the Petersson scalar product of two arbitrary cusp forms f and g as a combination of products of periods of opposity parity of f and g, since it then follows that any cusp form in the kernel of either r^+ or r^-has scalar product with itself equal to zero and hence vanishes. Such a formula was given by Haberland. The proof is elementary (essentially an application of Stokes's theorem applied to a cleverly chosen function in a fundamental domain for Γ), but a little long, so we will not give it, simply contenting ourselves with stating the result:

$$(f,g)=\frac{1}{3\,(2i)^{k-1}}\sum_{\substack{m,n\ge 0\\ m+n\le k-2\\ m\not\equiv n \pmod 2}}\frac{(k-2)!}{m!\,n!\,(k-2-m-n)!}\,r_m(f)\,\overline{r_n(g)}\,. \tag{11}$$

This formula can be extended to the case when one of f or g is an Eisenstein series (in which case the scalar product (f,g) vanishes) and then leads to a description of the 1-codimensional subspace $r^-(S_k)$ of W_k^-, namely

$$\mathrm{Im}(r_k^-)=\Big\{\sum_{n=0}^{k-2}c_nX^n\in W_k^-\mid\sum_{0\le r<n\le k-2}\frac{B_r}{r!}\,\frac{B_{k-r}}{k-r}\,\frac{n!}{(n-r+1)!}\,c_n=0\Big\}\,, \tag{12}$$

where B_r denotes the rth Bernoulli number.

5.2. Hecke operators and algebraicity results. The fact that the space W_{12}^+ is one-dimensional and spanned by the polynomial $4X^9-25X^7+42X^5-25X^3+4X$ shows that the "+"-period polynomial of the cusp form $\Delta\in S_{12}$ is given by

$$r^+(\Delta)(X)=\omega^+(\Delta)\,(4X^9-25X^7+42X^5-25X^3+4X)$$

for some complex number $\omega^+(\Delta)$ (in fact $\omega^+(\Delta) = 0{,}00926927\dots$). The same argument does not work for $r^-(\Delta)$, since W_{12}^- is two-dimensional, but in fact $r^-(\Delta)$ is also a multiple of polynomial with rational coefficients, namely

$$r^-(\Delta)(X) = \omega^-(\Delta)\,(\frac{36}{691}X^{10} - X^8 + 3X^6 - 3X^4 + X^2 - \frac{36}{691})$$

with $\omega^-(\Delta) = 0{,}114379\dots\, i$. Moreover, $\omega^+(\Delta)$ and $\omega^-(\Delta)$ are related to the Petersson scalar product (Δ, Δ) by $\omega^+(\Delta)\omega^-(\Delta) = 2^{10}\, i\,(\Delta, \Delta)$. These formulas are examples of the following theorem, which is the algebraicity result alluded to in the introduction to the chapter.

Theorem (Manin). *Let $f \in S_k$ be a Hecke eigenform. Then there exist two complex numbers $\omega^+(f) \in \mathbb{R}$, $\omega^-(f) \in i\mathbb{R}$ such that the polynomials $\omega^(f)^{-1} r^{\mp}(f)(X)$ have coefficients in the field K_f generated by the eigenvalues (= Fourier coefficients) of f. Moreover, $\omega^+(f)\omega^-(f)/i(f,f)$ also belongs to K_f (or even to $\mathbb{Q}$ if $\omega^{\pm}(f)$ are chosen suitably).*

Note that the last statement follows immediately from (11), so that the main assertion is the fact that both period polynomials $r^{\pm}(f)(X)$ are multiples of polynomials with coefficients in K_f. We will sketch two proofs of this assertion.

The first proof relies on the Rankin-Selberg method as treated in the last chapter. We wish to show that the product $r_m(f)r_n(f)$ of two periods of f with m and n of opposite parity is the product of $i(f,f)$ with a number belonging to K_f. By (2), this is equivalent to showing that $L^*(f, m+1)L^*(f, n+1)$ is a K_f-multiple of (f,f). Using the functional equation, we can assume that $m+1 = h+\nu$, $n+1 = h+l+\nu-1$ for some integers h, $l > 0$, $\nu \geq 0$, $h + l + 2\nu = k$, and h and l even. The desired statement now follows from formula (4) of 4.3 since $F_\nu(G_h, G_l)$ has rational Fourier coefficients (because G_h and G_l do) and the scalar product with an eigenform f of any modular form g with rational coefficients is automatically a K_f-multiple of (f,f) (since $(g,f)/(f,f)$ is just the coefficient of f in the eigenform decomposition of g). What's more, we get not only algebraicity results but explicit formulas: for instance, $F_2(G_4, G_4) = (2\pi i)^{-2}\big(20G_4''G_4 - 25{G_4'}^2\big)$ equals $\frac{1}{12}\Delta$, so eq. (4) of 4.3 gives

$$\frac{1}{12}\,(\Delta,\,\Delta) = \frac{1}{2^{11}}\binom{10}{4}\,L^*(\Delta, 9)\,L^*(\Delta, 6) = \frac{105\,i}{2^{11}}\,r_5(\Delta)\,r_8(\Delta)\,,$$

agreeing with the values $-\binom{10}{5}r_5(\Delta) = 42\omega^+(\Delta)$, $\binom{10}{2}r_8(\Delta) = \omega^-(\Delta)$ and $\omega^+(\Delta)\omega^-(\Delta) = 2^{10}i(\Delta,\Delta)$ given above.

The second approach, which also leads to explicit formulas for the period polynomials, is based on the action of Hecke operators. For each integer $n > 0$ denote by $\mathcal{M}_n$ the set of finite integral linear combinations of 2×2 matrices with integer coefficients and determinant n. We let $\mathcal{M}_n$ act on V_k by defining the action of $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ $(ad - bc = n)$ to be $\phi(X) \mapsto (cX+d)^{k-2}\phi((aX+b)/(cX+d))$.

Theorem. *For each natural number n define an element $\widetilde{T}_n \in \mathcal{M}_n$ as $\sum \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, where the sum ranges over all 2×2 matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ of determinant n satisfying the conditions $a > |c|$, $d > |b|$, $bc \le 0$ and also $-a < 2c \le a$ if $b = 0$, $-d < 2b \le d$ if $c = 0$. Then the spaces W_k^+ and W_k^- are stable under $\widetilde{T}_n$ for all k and all n, and the period polynomials of f and $T_n f$ for $f \in S_k$ are related by $r(T_n f) = r(f)|\widetilde{T}_n$. In particular, the polynomials $r^{\pm}(f)$ for a Hecke eigenform f are multiples of eigenvectors for the action of $\widetilde{T}_n$ on $W_k^{\pm}$.*

For example, we have

$$\widetilde{T}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \widetilde{T}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix},$$
$$\widetilde{T}_3 = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ 0 & 3 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ -1 & 1 \end{pmatrix}.$$

Using this and the calculation W_{12}^+ given in the last section, we find that (for instance) $\widetilde{T}_2$ acts on the one-dimensional space W_{12}^+ with the unique eigenvalue -24 and on the two-dimensional space W_{12}^- with the two eigenvalues $+2049$ and -24, the corresponding eigenvectors being $X^{10} - 1$ and $\frac{36}{691}(X^{10} - 1) - X^2(X^2 - 1)^3$. This shows that the unique cusp form of weight 12 has eigenvalue -24 under T_2 and $+$- and $-$-period polynomials proportional to $4X^9 - 25X^7 + 42X^5 - 25X^3 + 4X$ and $\frac{36}{691}(X^{10} - 1) - X^2(X^2 - 1)^3$, respectively. (The other eigenvalue $+2049$ on W_{12}^- comes from the Eisenstein series; cf. Exercise 5).

We do not give the proof of the theorem, but only make a few remarks about how it goes. Clearly the abelian group $\mathcal{M}_n$ is a left and right module over the group ring $\mathbb{Z}[\Gamma]$ introduced in the last section. The crucial fact about $\widetilde{T}_n$ is the congruence

$$(13) \qquad (1 - S)\,\widetilde{T}_n = T_n^{\infty}\,(1 - S) \qquad (\mathrm{mod}\ (T - 1)\,\mathcal{M}_n),$$

where $T = US = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ denotes translation by 1 and T_n^{∞} is the element $\sum \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ (sum over a, b, d with $ad = n$ and $0 \le b < d$) of $\mathcal{M}_n$. On the other hand, it can be shown without difficulty that the period polynomial $r(f)$ of an arbitrary cusp form $f = \sum a_n\, q^n \in S_k$ is up to a simple factor equal to $\widetilde{f}|_{2-k}(1 - S)$, where $\widetilde{f}(\tau) = \sum n^{1-k}\, a_n\, q^n$ is the $(k-1)$-fold integral of f (so-called "Eichler integral"). The corresponding Eichler integral for $T_n f$ is just $T_n^{\infty}\widetilde{f}$, and this fact together with (13) and $\widetilde{f}|(T - 1) = 0$ easily leads to a proof of the identity claimed.

Exercises

1. Prove the invariance property (6) of the scalar product on V_k defined by (5). (Hint: Show that (5) is equivalent to $(\phi, \psi_t) = \phi(t)$ for any $t \in \mathbb{C}$, where $\psi_t(X) = (X - t)^{k-2}$, and then prove (6) in the special case $\psi = \psi_t$.)

2. Prove the dimension formulas (10). (Hint: For the dimension of C_k, there are two ways. One is to compute the action of U on the basis $\{(X + \omega)^n (X + \omega^2)^m\}_{n+m=k-2}$ of V_k and show that C_k is spanned by the elements with $n \equiv m \pmod 3$. The other is to show

that $\mathrm{tr}(U) = \mathrm{tr}(U^2) = \dim C_k - \dim D_k$ and show by considering the action of U on the basis $\{X^n\}_{0 \le n \le k-2}$ of V_k that $\mathrm{tr}(U) = t_k := \sum_{n+m=k-2} (-1)^n \binom{m}{n}$; this is then shown by considering the generating function identity $\sum_{k \ge 2} t_k x^{k-2} = (1 - x(1-x))^{-1}$ to be ± 1 or 0 according to the value of k modulo 6.)

3. Let D be a positive non-square integer congruent to 0 or 1 modulo 4 and n a natural number. Show that the function

$$P_{n,D}(X) = \sum_{\substack{a,b,c \in \mathbb{Z} \\ a > 0 > c \\ b^2 - 4ac = D}} (aX^2 + bX + c)^{2n-1}$$

belongs to W_{4n}^-. (In particular, $P_{n,D}(X) + \kappa(X^{4n-2} - 1)$ equals $r^-(f)$ for some cusp form $f \in S_{4n}$ and number $\kappa = \kappa_{n,D}$ by the Eichler-Shimura theorem; f and κ can be calculated explicitly and κ turns out to be essentially the value at $s = 2n$ of the Dedekind zeta function of $\mathbb{Q}(\sqrt{D})$, and then (12) gives an elementary formula for this value in terms of the solutions of $b^2 - 4ac = D$ with $ac < 0$.)

4. Verify formula (13), either for all n or else just for $n = 1$, 2, and 3.

5. Verify that the element $X^{k-2} - 1$ of W_k^- is an eigenvalue of $\widetilde{T}_n$ with eigenvalue $\sigma_{k-1}(n)$ for all n.

Fix an even weight $k > 0$ and let

$$t(n) = t_k(n) = \mathrm{Tr}(T(n), M_k), \qquad t^0(n) = t_k^0(n) = \mathrm{Tr}(T(n), S_k)$$

denote the traces of the nth Hecke operator $T(n)$ on the spaces of modular forms and cusp forms, respectively, of weight k. If we choose as a basis for M_k or S_k a $\mathbb{Z}$-basis of the lattice of forms having integral Fourier coefficients (which we know we can do by the results of Chapter 1), then the matrix representing the action of $T(n)$ with respect to this basis also has integral coefficients. Hence $t(n)$ and $t^0(n)$ are integers. The splitting $M_k = S_k \bigoplus \langle G_k \rangle$ and the formula $T_n(G_k) = \sigma_{k-1}(n) G_k$ for $k > 2$ imply

$$(1) \qquad t_k(n) = t_k^0(n) + \sigma_{k-1}(n) \qquad (n \geq 1,\ k > 2).$$

The importance of knowing $t^0(n)$ is as follows. Let $\mathbf{t}^0(\tau) = \mathbf{t}_k^0(\tau) = \sum_{n=1}^{\infty} t_k^0(n) q^n$. Then $\mathbf{t}^0$ is itself a cusp form of weight k on Γ_1 and its images under all Hecke operators (indeed, under $T(n_1), \ldots, T(n_d)$ for any $\{n_j\}_{j=1}^{d=\dim S_k}$ for which the n_1st, …, n_dth Fourier coefficients of forms in S_k are linearly independent) generate the space S_k. To see this, let $f_i(\tau) = \sum_{n>0} a_i(n) q^n$ $(1 \leq i \leq d)$ be the Hecke forms in S_k. We know that they form a basis and that the action of $T(n)$ on this basis is given by the diagonal matrix $\mathrm{diag}(a_1(n), \ldots, a_d(n))$. Hence the trace $t^0(n)$ equals $a_1(n) + \ldots + a_d(n)$ and $\mathbf{t}_k^0$ is just $f_1 + \ldots + f_d$, which is indeed in S_k; the linear independence of the f_i and the fact that the matrix $(a_i(n_j))_{1 \leq i,j \leq d}$ is invertible then imply that the d forms $T(n_j)(\mathbf{t}^0) = \sum_{i=1}^{d} a_i(n_j) f_i$ are linearly independent and hence span S_k as claimed. Having a formula for $\mathrm{Tr}(T(n))$ thus gives an algorithm for obtaining all cusp forms of a given weight (and level).

In this chapter we will state and prove such a formula. The basic tool we will use is the holomorphic version of the Rankin-Selberg method described in Chapter 4, but now in the case when the Eisenstein series G_h and the modular form g have half-integral weight. We therefore begin by giving a brief discussion of forms of half-integral weight in 6.1. A technical complication is that the specific Eisenstein series we will need (weight $h = 3/2$) is not quite holomorphic. To correct for this we require another important tool in the theory of modular forms, the so-called holomorphic projection operator. This will be treated in 6.2. The statement and proof of the trace formula are given in 6.3.

6.1. Eisenstein series of half-integral weight. In Chapter 3, there was no reason to look only at quadratic forms in an even number of variables. If we take the simplest possible quadratic form $Q(x_1) = x_1^2$, then the associated theta-series

$$\theta(\tau) = \sum_{n \in \mathbb{Z}} q^{n^2} = 1 + 2q + 2q^4 + 2q^9 + \cdots$$

is the square-root of the first example in that section and as such satisfies the transformation equation

$$\theta\left(\frac{a\tau+b}{c\tau+d}\right) = \epsilon(c\tau+d)^{\frac{1}{2}}\theta(\tau) \quad \forall\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(4)$$

for a certain number $\epsilon = \epsilon_{c,d}$ satisfying $\epsilon^4 = 1$ (ϵ can be given explicitly in terms of the Kronecker symbol $\left(\frac{c}{d}\right)$). We say that θ is a **modular form of weight** $\frac{1}{2}$. More generally, we can define modular forms of any half-integral weight $r+\frac{1}{2}$ ($r \in \mathbb{N}$). A particularly convenient space of such forms, analogous to the space M_k of integral-weight modular forms on the full modular group, is the space $M_{r+\frac{1}{2}}$ introduced by W. Kohnen. It consists of all f satisfying the transformation law $f\left(\frac{a\tau+b}{c\tau+d}\right) = \left(\epsilon_{c,d}(c\tau+d)^{\frac{1}{2}}\right)^{2r+1} f(\tau)$ for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(4)$ (equivalently, f/θ^{2r+1} should be $\Gamma_0(4)$-invariant) and having a Fourier expansion of the form $\sum_{n\geq 0} a(n)q^n$ with $a(n) = 0$ whenever $(-1)^r n$ is congruent to 2 or 3 modulo 4. For $r \geq 2$ this space contains an Eisenstein series $G_{r+\frac{1}{2}}$ calculated by H. Cohen. We do not give the definition and the calculation of the Fourier expansion of these Eisenstein series, which are similar in principle but considerably more complicated than in the integral weight case. Unlike the case of integral weight, where the Fourier coefficients were elementary arithmetic functions, the Fourier coefficients now turn out to be number-theoretical functions of considerable interest. Specifically, we have

$$G_{r+\frac{1}{2}}(\tau) = \sum_{\substack{n=0 \\ (-1)^r n \equiv 0 \text{ or } 1 \pmod 4}}^{\infty} H(r,n)\, q^n$$

where $H(r,n)$ is a special value of some L-series, e.g. $H(r,0) = \zeta(1-2r) = -\frac{B_{2r}}{2r}$ (where $\zeta(s)$ is the Riemann zeta-function and B_m the mth Bernoulli number), $H(r,1) = \zeta(1-r)$, and more generally $H(r,n) = L_\Delta(1-r)$ if the number $\Delta = (-1)^r n$ is equal to either 1 or the discriminant of a real or imaginary quadratic field, where the L-series $L_\Delta(s)$ is defined as the analytic continuation of the Dirichlet series $\sum_{n=1}^{\infty}\left(\frac{\Delta}{n}\right)n^{-s}$. These numbers are known to be rational, with a bounded denominator for a fixed value of r. The first few cases are

$$\begin{aligned}
G_{2\frac{1}{2}}(\tau) &= \tfrac{1}{120} - \tfrac{1}{12}q - \tfrac{7}{12}q^4 - \tfrac{3}{5}q^5 - q^8 - \tfrac{25}{12}q^9 - 2q^{12} - 2q^{13} - \tfrac{55}{12}q^{16} - 4q^{17} - \cdots \\
G_{3\frac{1}{2}}(\tau) &= -\tfrac{1}{252} - \tfrac{2}{9}q^3 - \tfrac{1}{2}q^4 - \tfrac{16}{7}q^7 - 3q^8 - 6q^{11} - \tfrac{74}{9}q^{12} - 16q^{15} - \tfrac{33}{2}q^{16} - \cdots \\
G_{4\frac{1}{2}}(\tau) &= \tfrac{1}{240} + \tfrac{1}{120}q + \tfrac{121}{120}q^4 + 2q^5 + 11q^8 + \tfrac{2161}{120}q^9 + 46q^{12} + 58q^{13} + \cdots \\
G_{5\frac{1}{2}}(\tau) &= -\tfrac{1}{132} + \tfrac{1}{3}q^3 + \tfrac{5}{2}q^4 + 32q^7 + 57q^8 + \tfrac{2550}{11}q^{11} + \tfrac{529}{3}q^{12} + 992q^{15} + \cdots .
\end{aligned}$$

In each of these four cases, the space $M_{r+\frac{1}{2}}$ is one-dimensional, generated by $G_{r+\frac{1}{2}}$; in general, $M_{r+\frac{1}{2}}$ has the same dimension as M_{2r}.

Just as the case of G_2, the Fourier expansion of $G_{r+1/2}$ still makes sense for $r=1$, but the analytic function it defines is no longer a modular form. Specifically, the function $H(r,n)$

when $r = 1$ is equal to the **Hurwitz-Kronecker class number** $H(n)$, defined for $n > 0$ as the number of $SL_2(\mathbb{Z})$-equivalence classes of binary quadratic forms of discriminant $-n$, each form being counted with a multiplicity equal to 1 divided by the order of its stabilizer in $SL_2(\mathbb{Z})$ (this order is 2 for a single equivalence class of forms if n is 4 times a square, 3 for a single class if n is 3 times a square, and 1 in all other cases). Thus the form $G_{3/2} = \sum_n H(n)q^n$ has a Fourier expansion beginning

$$G_{\frac{3}{2}}(\tau) = -\frac{1}{12} + \frac{1}{3}q^3 + \frac{1}{2}q^4 + q^7 + q^8 + q^{11} + \frac{4}{3}q^{12} + 2q^{15} + \frac{3}{2}q^{16} + q^{19} + 2q^{20} + 3q^{23} + \cdots .$$

As with G_2 we can use "Hecke's trick" to define a function $G^*_{3/2}$ which is not holomorphic but transforms like a holomorphic modular form of weight 3/2. The Fourier expansion of this non-holomorphic modular form differs from that of $G_{3/2}$ only at negative square exponents:

$$G^*_{\frac{3}{2}}(\tau) = \sum_{n=0}^{\infty} H(n)\, q^n + \frac{1}{16\pi\sqrt{v}} \sum_{f\in\mathbb{Z}} \beta(4\pi f^2 v)\, q^{-f^2} \tag{2}$$

where v denotes the imaginary part of τ and $\beta(t)$ the function $\int_1^\infty x^{-3/2}\, e^{-xt}\, dx$, which can be expressed in terms of the error function.

6.2. Holomorphic projection. We know that S_k has a scalar product $(\cdot,\cdot)$ which is non-degenerate (since $(f, f) > 0$ for every $f \neq 0$ in S_k). It follows that any linear functional $L : S_k \to \mathbb{C}$ can be represented as $f \mapsto (f, \phi_L)$ for a unique cusp form $\phi_L \in S_k$.

Now suppose that $\Phi : \mathfrak{H} \to \mathbb{C}$ is a function which is not necessarily holomorphic but transforms like a holomorphic modular form of weight k, and that $\Phi(\tau)$ has reasonable (say, at most polynomial) growth in $v = \Im(\tau)$ as $v \to \infty$. Then the scalar product $(f, \Phi) = \iint_{\mathfrak{H}/\Gamma_1} v^k f(\tau)\overline{\Phi(\tau)}\, d\mu$ converges for every f in S_k, and since $f \mapsto (f, \Phi)$ is linear, there exists a unique function $\phi \in S_k$ satisfying $(f, \phi) = (f, \Phi)$ for every $f \in S_k$. Clearly $\phi = \Phi$ if Φ is already in S_k, so that the operator π_{hol} which assigns ϕ to Φ is a projection from the infinite dimensional space of functions in $\mathfrak{H}$ transforming like modular forms of weight k to the finite dimensional subspace of holomorphic cusp forms. In this section we derive a formula for the Fourier coefficients of $\pi_{\text{hol}}(\Phi)$.

Let $\sum_{n\in\mathbb{Z}} c_n(v)e^{2\pi i n u}$ denote the Fourier development of $\Phi(\tau)$ and $\sum_{n=1}^\infty c_n q^n$ that of its holomorphic projection ϕ. We apply the basic identity $(\phi, f) = (\Phi, f)$ to the Poincaré series $f = P_n \in S_k$ defined in Exercise 3 of Chapter 4. By that exercise, we know that the scalar product of ϕ with P_n is $(k-2)!c_n/(4\pi n)^{k-1}$. The same unfolding argument used to prove that formula can be applied also to the non-holomorphic modular form Φ and gives

$$\begin{aligned}(\Phi, P_n) &= \int_0^\infty \left(\int_0^1 \Phi(u + iv)\, du\right) e^{-2\pi n v} v^{k-2}\, dv \\ &= \int_0^\infty c_n(v) e^{-2\pi n v} v^{k-2}\, dv\end{aligned}$$

provided that the interchange of summation and integration implicit in the first step is justified. This is certainly the case if the scalar product (Φ, P_n) remains convergent after replacing Φ by $|\Phi|$ and P_n by its majorant $\hat{P}_n(\tau) = \sum_{\Gamma_\infty\backslash\Gamma_1} |(c\tau+d)^{-k} e^{2\pi i n\gamma(\tau)}|$. We have

$$\begin{aligned}\hat{P}_n(\tau) &< |e^{2\pi i n\tau}| + \sum_{c\neq 0}\sum_{(d,c)=1} |c\tau+d|^{-k} \\ &= e^{-2\pi n v} + \frac{1}{\zeta(k)} v^{-k/2}\left[G(\tau, \frac{k}{2}) - \zeta(k)v^{k/2}\right],\end{aligned}$$

with $G(\tau, \frac{k}{2})$ the non-holomorphic Eisenstein series introduced in §4.1. The expansions there show that $G(\tau, \frac{k}{2}) - \zeta(k)v^{k/2} = \mathrm{O}(v^{1-k/2})$ as $v \to \infty$, so $\hat{P}_n(\tau) = \mathrm{O}(v^{1-k})$. The convergence of $\iint_{\mathfrak{H}/\Gamma_1} |\Phi|\hat{P}_n v^{k-2}\,du\,dv$ is thus assured if $\Phi(\tau)$ decays like $\mathrm{O}(v^{-\epsilon})$ as $v \to \infty$ for some positive number ϵ. Finally, we can weaken the condition $\Phi(\tau) = \mathrm{O}(v^{-\epsilon})$ to $\Phi(\tau) = c_0 + \mathrm{O}(v^{-\epsilon})$ ($c_0 \in \mathbb{C}$) by the simple expedient of subtracting $c_0\frac{-2k}{B_k}G_k(\tau)$ from $\Phi(\tau)$ and observing that G_k is orthogonal to cusp forms by the same calculation as above with $n = 0$ (G_k is proportional to P_0). We have thus proved the following result, first stated by J. Sturm under slightly different hypotheses:

Holomorphic Projection Lemma. *Let $\Phi : \mathfrak{H} \to \mathbb{C}$ be a continuous function satisfying*

(i) $\Phi(\gamma(\tau)) = (c\tau+d)^k\Phi(\tau)$ *for all* $\gamma = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \Gamma_1$ *and* $\tau \in \mathfrak{H}$; *and*

(ii) $\Phi(\tau) = c_0 + O(v^{-\epsilon})$ *as* $v = \Im(\tau) \to \infty$,

for some integer $k > 2$ and numbers $c_0 \in \mathbb{C}$ and $\epsilon > 0$. Then the function $\phi(\tau) = \sum_{n=0}^{\infty} c_n q^n$ with $c_n = \frac{(4\pi n)^{k-1}}{(k-2)!}\int_0^\infty c_n(v)e^{-2\pi n v}v^{k-2}\,dv$ for $n > 0$ belongs to M_k and satisfies $(f, \phi) = (f, \Phi)$ for all $f \in S_k$.

As an example, take $\Phi = (G_2^*)^2$, where G_2^* is the non-holomorphic Eisenstein series of weight 2 introduced in Chapter 1. Using the Fourier expansion $G_2^* = \frac{1}{8\pi v} + G_2 = \frac{1}{8\pi v} - \frac{1}{24} + \sum_1^\infty \sigma_1(n)q^n$ given there, we find

$$\begin{aligned}\Phi(\tau) =& \Big(\frac{1}{576} - \frac{1}{96\pi v} + \frac{1}{64\pi^2v^2}\Big) \\ &+ \sum_{n=1}^{\infty}\left(-\frac{1}{12}\sigma_1(n) + \sum_{m=1}^{n-1}\sigma_1(m)\sigma_1(n-m) + \frac{1}{4\pi v}\sigma_1(n)\right)q^n,\end{aligned}$$

so that the hypotheses of the holomorphic projection lemma are satisfied with $k = 4$, $c_0 = \frac{1}{576}$, $\epsilon = 1$ and $c_n(v) = \big(-\frac{1}{12}\sigma_1(n) + \sum_{m=1}^{n-1}\sigma_1(m)\sigma_1(n-m) + \frac{1}{4\pi v}\sigma_1(n)\big)e^{-2\pi n v}$. The lemma then gives $\sum c_nq^n \in M_4$ with $c_n = -\frac{1}{12}\sigma_1(n) + \sum_{m=1}^{n-1}\sigma_1(m)\sigma_1(n-m) + \frac{1}{2}n\sigma_1(n)$ for $n \geq 1$. Since $\sum_0^\infty c_nq^n \in M_4 = \langle G_4\rangle$, we must have $c_n = 240c_0\sigma_3(n) = \frac{5}{12}\sigma_3(n)$ for all $n > 0$, an identity that the reader can check for small values of n.

Similarly, if $f = \sum_0^\infty a_nq^n$ is a modular form of weight $l \geq 4$, then $\Phi = fG_2^*$ satisfies the hypotheses of the lemma with $k = l+2$, $c_0 = -\frac{1}{24}a_0$ and $\epsilon = 1$, and we find that $\pi_{\mathrm{hol}}(fG_2^*) = fG_2 + \frac{1}{4\pi i l}f' \in M_{l+2}$.

6.3. The Eichler-Selberg trace formula. Let $k>0$ be even and $t_k(n)$, $t_k^0(n)$ be defined as in the beginning of the chapter.

Theorem (Eichler, Selberg). *Let $H(N)$ $(N \geq 0)$ be the Kronecker-Hurwitz class numbers defined in Section 6.1 and denote by $p_k(t,n)$ the homogeneous polynomial*

$$p_k(t,n) = \sum_{0 \leq r \leq \frac{k}{2}-1} \binom{k-2-r}{r} (-n)^r t^{k-2-2r} = \mathit{Coeff}_{X^{k-2}}\Bigl(\frac{1}{1-tX+nX^2}\Bigr)$$

of degree $\frac{k}{2}-1$ in t^2 and n (thus $p_2(t,n)=1$, $p_4(t,n)=t^2-n$, $p_6(t,n)=t^4-3t^2n+n^2$, etc.). Then

$$t_k(n) = -\frac{1}{2} \sum_{t \in \mathbb{Z}, t^2 \leq 4n} p_k(t,n) H(4n-t^2) + \frac{1}{2} \sum_{d|n} \max\{d, n/d\}^{k-1} \qquad (k \geq 2),$$

$$t_k^0(n) = -\frac{1}{2} \sum_{t \in \mathbb{Z}, t^2 \leq 4n} p_k(t,n) H(4n-t^2) - \frac{1}{2} \sum_{d|n} \min\{d, n/d\}^{k-1} \qquad (k \geq 4).$$

There is an analogous trace formula for forms of higher level (say, for the trace of $T(n)$ on $M_k(\Gamma_0(N))$ for n and N coprime), but the statement is more complicated and we omit it. The equivalence of the two formulas in the theorem (for $k>2$) follows from (1), since

$$\frac{1}{2} \sum_{d|n} \bigl(\min\{d,n/d\}^{k-1} + \max\{d,n/d\}^{k-1}\bigr) = \frac{1}{2} \sum_{d|n} \bigl(d^{k-1} + (n/d)^{k-1}\bigr) = \sigma_{k-1}(n).$$

Note also that $t_2(n)=0$ and $t_k^0(n)=0$ for $k \in \{2,4,6,8,10,14\}$ and all n, since the spaces M_2 and S_k are 0-dimensional in these cases. Equating to zero the expressions for $t_2(n)$ and $t_4^0(n)$ given in the theorem gives two formulas of the form

$$H(4n) + 2H(4n-1) + \ldots = 0, \qquad -nH(4n) - 2(n-1)H(4n-1) + \ldots = 0,$$

where the terms "$\ldots$" involve only $H(4m)$ and $H(4m-1)$ with $m<n$. Together, these formulas give a rapid inductive method of computing all the Kronecker-Hurwitz class numbers $H(N)$.

We now sketch a proof of the Eichler-Selberg trace formula, using the "holomorphic version" of the Rankin-Selberg method proved in 4.3, but applied in the case when the Eisenstein series G_h and the modular form g have half-integral weight. The basic identities (1) and (2) of 4.3 remain true in this context with slight modifications due to the fact that the functions G_h and g are modular forms on $\Gamma_0(4)$ rather than $SL_2(\mathbb{Z})$. They can be avoided by using the operator U_4 which maps a periodic function $h(\tau)$ to $\frac{1}{4}\sum_{j=0}^{3} h(\frac{1}{4}(\tau+j))$ (or in terms of Fourier series, $U_4 : \sum c_n q^n \mapsto \sum c_{4n} q^n$). One can check that $U_4(F_\nu(G_h,g))$ belongs to $M_k(\Gamma_1)$ if h and l are half-integral, $k=h+l+2\nu$ is even, and g belongs to the space M_l defined in 6.1. In this situation, formula (2) of 4.3 still holds except for the values of the constant factors occurring. In particular, if $h = r + \frac{1}{2}$ with r odd and we take for g the basic theta-series $\theta(\tau) = 1 + 2\sum_{n=1}^{\infty} q^{n^2}$ of weight $\frac{1}{2}$ on $\Gamma_0(4)$, then we find

$$(3) \qquad (f, U_4(F_\nu(G_{r+\frac{1}{2}}, \theta))) = c_{\nu,r} \sum_{n=1}^{\infty} \frac{a(n^2)}{(n^2)^{k-\nu-1}} \qquad (r > 1 \text{ odd}, \ \nu \geq 0, \ k = r + 2\nu + 1),$$

where $c_{\nu,r}$ is an explicitly known constant depending only on r and ν. We want to apply this formula in the case $r = 1$. Here the function $G_{3/2}$ is not a modular form and must be replaced by the function $G_{3/2}^*$ defined by (2). The function $U_4(F_\nu(G_{3/2}^*, \theta))$ is no longer holomorphic, but we can apply the holomorphic projection operator of §6.2 to replace it by a holomorphic modular form without changing its Petersson scalar product with the holomorphic cusp form f. Moreover, for $r = 1$ we have $\nu = \frac{1}{2}(k - r - 1) = \frac{k}{2} - 1$ and hence $2(k - \nu - 1) = k$, so the right-hand side of (3) is proportional to $\sum \frac{a(n^2)}{n^s}\big|_{s=k}$ and hence, by formula (5) of Section 4.2, to (f, f) if f is a normalized Hecke eigenform. Thus finally

$$(f, \pi_{\mathrm{hol}}(U_4(F_\nu(G_{3/2}^*, \theta)))) = c_k \, (f, f)$$

for all Hecke forms $f \in S_k$, where c_k depends only on k (in fact, $c_k = -2^{k-1}\binom{\nu - \frac{1}{2}}{\nu}$). But since $\mathbf{t}^0(\tau)$ is the sum of all such eigenforms, and since distinct eigenforms are orthogonal, we also have $(f, \mathbf{t}^0) = (f, f)$ for all Hecke forms. It follows that

$$(4) \qquad \pi_{\mathrm{hol}}(U_4(F_\nu(G_{3/2}^*, \theta))) = c_k \, \mathbf{t}^0(\tau) + c_k' G_k(\tau)$$

for some constant c_k'.

It remains only to compute the Fourier expansion of the function on the left of (4). We have

$$\theta(\tau) = \sum_{t \in \mathbb{Z}} q^{t^2}, \qquad G_{\frac{3}{2}}(\tau) = \sum_{m=0}^{\infty} H(m) q^m$$

and hence

$$\begin{aligned} F_\nu(\theta(\tau), G_{\frac{3}{2}}(\tau)) &= (2\pi i)^{-\nu} \sum_{\mu=0}^{\nu} (-1)^\mu \binom{\nu - \frac{1}{2}}{\mu} \binom{\nu + \frac{1}{2}}{\nu - \mu} \theta^{(\nu-\mu)}(\tau) G_{\frac{3}{2}}^{(\mu)}(\tau) \\ &= \sum_{\substack{m, t \in \mathbb{Z} \\ m \geq 0}} \sum_{\mu=0}^{\nu} (-1)^\mu \binom{\nu - \frac{1}{2}}{\mu} \binom{\nu + \frac{1}{2}}{\nu - \mu} t^{2(\nu-\mu)} m^\mu H(m) q^{m+t^2}, \end{aligned}$$

so

$$\begin{aligned} &U_4(F_\nu(\theta(\tau), G_{\frac{3}{2}}(\tau))) \\ &\quad = \sum_{n=0}^{\infty} \sum_{t^2 \leq 4n} \sum_{\mu=0}^{\nu} (-1)^\mu \binom{\nu - \frac{1}{2}}{\mu} \binom{\nu + \frac{1}{2}}{\nu - \mu} t^{2\nu - 2\mu} (4n - t^2)^\mu H(4n - t^2) q^n \\ &\quad = -\frac{1}{2} c_k \sum_{n=0}^{\infty} \sum_{t^2 \leq 4n} p_k(t, n) H(4n - t^2) q^n \end{aligned}$$

(recall that $k = 2\nu + 2$). On the other hand, the difference of $G^*_{3/2}$ and $G_{3/2}$ is a linear combination of terms q^{-f^2} with coefficients which are analytic functions of $v = \Im(\tau)$. Hence the coefficient of q^n in $U_4(F_\nu(\theta, G^*_{3/2} - G_{3/2}))$ is a sum over all pairs $(t, f) \in \mathbb{Z}^2$ with $t^2 - f^2 = 4n$ of a certain analytic function of v. Applying π_{hol} means that this expression must be multiplied by $v^{k-2}e^{-4\pi nv}$ and integrated from $v = 0$ to $v = \infty$. The integral turns out to be elementary and one finds after a little calculation

$$\text{coefficient if } q^n \text{ in } \pi_{\mathrm{hol}}\big(U_4(F_\nu(\theta, G^*_{3/2} - G_{3/2}))\big)$$
$$= -\frac{1}{4}c_k \sum_{\substack{t,f\in\mathbb{Z}\\ t^2-f^2=4n}} \left(\frac{|t|-|f|}{2}\right)^{k-1} = -\frac{1}{2}c_k \sum_{\substack{d|n\\ d>0}} \min(d, \frac{n}{d})^{k-1}.$$

Adding this to the preceding formula, and comparing with (4), we find that the constant c'_k in (4) must be 0 and that we have obtained the result stated in the theorem.

Exercises

1. Verify the examples of holomorphic projections given at the end of 6.2. More generally, compute $\pi_{\mathrm{hol}}(F_\nu(f, G^*_2))$ for any holomorphic modular form $f \in M_l$ (note that the derivative $d^\mu/d\tau^\mu$ in the definition of F_ν must be replaced by $\partial^\mu/\partial\tau^\mu$ when F_ν is applied to a non-holomorphic form).

2. Use the recursions given by the cases $k = 2$ and $k = 4$ of the trace formula to compute $H(N)$ for $0 \le N \le 20$ and then use the case $k = 12$ to calculate the first few coefficients of the discriminant function $\Delta \in S_{12}$.

3. Compute the contribution of the terms $(16\pi\sqrt{v})^{-1}\beta(4\pi f^2 v)\, q^{-f^2}$ of $G^*_{3/2}$ and q^{t^2} of θ to $F_{k-2}(\theta, G^*_{3/2})$, at least for small k (say $k = 2$ and $k = 4$) and evaluate the contribution to $\pi_{\mathrm{hol}}(U_4(F_\nu(\theta, G^*_{3/2})))$ obtained by multiplying this by $v^{k-2}e^{-4\pi nv}$ $(n = (t^2 - f^2)/4)$ and integrating over $0 \le v \le \infty$.

A1. The Poisson summation formula. This is the identity

$$\sum_{n\in\mathbb{Z}} \varphi(x+n) \;=\; \sum_{r\in\mathbb{Z}} \left(\int_{\mathbb{R}} \varphi(t)e^{-2\pi irt}\,dt\right) e^{2\pi irx}, \tag{1}$$

where $\varphi(x)$ is any continuous function on $\mathbb{R}$ which decreases rapidly (say, at least like $|x|^{-c}$ with $c>1$) as $x\to\infty$. The proof is simple: the growth condition on φ ensures that the sum on the left-hand side converges absolutely and defines a continuous function $\Phi(x)$. Clearly $\Phi(x+1)=\Phi(x)$, so Φ has a Fourier expansion $\sum_{r\in\mathbb{Z}} c_r e^{2\pi irx}$ with Fourier coefficients c_r given by $\int_0^1 \Phi(x)e^{-2\pi irx}\,dx$. Substituting into this formula the definition of Φ, we find

$$\begin{aligned} c_r &= \int_0^1 \left(\sum_{n=-\infty}^{\infty} \varphi(x+n)e^{-2\pi ir(x+n)}\right) dx \\ &= \sum_{n=-\infty}^{\infty} \int_n^{n+1} \varphi(x)e^{-2\pi irx}\,dx = \int_{-\infty}^{\infty} \varphi(x)e^{-2\pi irx}\,dx, \end{aligned}$$

as claimed. If we write $\hat\varphi(t)$ for the **Fourier transform** $\int_{-\infty}^{\infty}\varphi(x)e^{-2\pi itx}\,dx$ of φ, then (1) can be written in the form $\sum_n \varphi(x+n)=\sum_r \hat\varphi(r)e^{2\pi irx}$, where both summations are over $\mathbb{Z}$. The special case $x=0$ has the more symmetric form $\sum_n\varphi(n)=\sum_r\hat\varphi(r)$, which is actually no less general since replacing $\varphi(x)$ by $\varphi(x+a)$ replaces $\hat\varphi(t)$ by $\hat\varphi(t)e^{2\pi ita}$; it is in this form that the Poisson summation formula is often stated.

As a first application, we take $\varphi(x)=(x+iy)^{-k}$, where y is a positive number and k an integer ≥ 2. This gives the **Lipschitz formula**

$$\sum_{n\in\mathbb{Z}} \frac{1}{(z+n)^k} \;=\; \frac{(-2\pi i)^k}{(k-1)!}\sum_{r=1}^{\infty} r^{k-1}e^{2\pi irz} \qquad (z\in\mathfrak{H},\ k\in\mathbb{Z}_{\geq 2}),$$

which can also be proved by expanding the right hand side of Euler's identity

$$\sum_{n\in\mathbb{Z}} \frac{1}{z+n} \;=\; \frac{\pi}{\tan\pi z} \;=\; -\pi i - 2\pi i\,\frac{e^{2\pi iz}}{1-e^{2\pi iz}}$$

as a geometric series in $e^{2\pi iz}$ and differentiating $k-1$ times with respect to z.

As a second application, take $\varphi(x)=e^{-\pi ax^2}$ with $a>0$. Then $\hat\varphi(t)=a^{-\frac12}e^{-\pi t^2/a}$, so we get

$$\sum_{n=-\infty}^{\infty} e^{-\pi a(x+n)^2} = \sqrt{\frac{1}{a}}\sum_{r=-\infty}^{\infty} e^{-\pi a^{-1}r^2+2\pi irx} \qquad (x\in\mathbb{R}) \tag{2}$$

(the formula is actually valid for all $x \in \mathbb{C}$, as one sees by replacing $\varphi(x)$ by $\varphi(x+iy)$ with $y \in \mathbb{R}$). This identity, and its generalizations to higher-dimensional sums of Gaussian functions, is the basis of the theory of theta functions.

Finally, if s is a complex number of real part greater than 1, then taking $\varphi(x) = |x+iy|^{-s}$ with $y > 0$ leads to the following non-holomorphic generalization of the Lipschitz formula:

$$\sum_{n\in\mathbb{Z}} \frac{1}{|z+n|^s} = y^{1-s} \sum_{r=-\infty}^{\infty} k_{s/2}(2\pi r y) e^{2\pi i r x} \qquad (z = x+iy \in \mathfrak{H},\ \Re(s) > 1),$$

where $k_s(t) = \int_{-\infty}^{\infty} e^{-itx}(x^2+1)^{-s}\,dx$. The function $k_s(t)$ can be expressed in terms of the gamma function $\Gamma(s)$ and the **K-Bessel function** $K_\nu(t) = \int_0^\infty e^{-t\cosh u} \cosh(\nu u)\,du$ ($\nu \in \mathbb{C}$, $t > 0$) by

$$k_s(t) = \begin{cases} \frac{2\pi^{\frac12}}{\Gamma(s)} \left(\frac{|t|}{2}\right)^{s-\frac12} K_{s-\frac12}(|t|) & \text{if } t \neq 0, \\ \frac{\pi^{\frac12}\Gamma(s-\frac12)}{\Gamma(s)} & \text{if } t = 0 \end{cases}$$

(cf. Appendix A2), so, replacing s by $2s$, we can rewrite the result as

$$\sum_{n\in\mathbb{Z}} \frac{1}{|z+n|^{2s}} = \frac{\pi^{\frac12}\Gamma(s-\frac12)}{\Gamma(s)} y^{1-2s} + \frac{2\pi^s}{\Gamma(s)}\, y^{\frac12 - s} \sum_{r\neq 0} |r|^{s-\frac12} K_{s-\frac12}(2\pi|r|y)\, e^{2\pi i r x} \tag{3}$$
$$(z = x+iy \in \mathfrak{H},\ \Re(s) > \frac{1}{2}).$$

This formula is used for computing the Fourier development of the non-holomorphic Eisenstein series (cf. §5.3).

A2. The gamma function and the Mellin transform. The integral representation $n! = \int_0^\infty t^n e^{-t}\,dt$ is generalized by the definition of the **gamma function**

$$\Gamma(s) = \int_0^\infty t^{s-1}\, e^{-t}\,dt \qquad (s \in \mathbb{C},\ \Re(s) > 0). \tag{1}$$

Thus $n! = \Gamma(n+1)$ for n a nonnegative integer. Integration by parts gives the functional equation $\Gamma(s+1) = s\Gamma(s)$, generalizing the formula $(n+1)! = (n+1)n!$ and also permitting one to define the Γ-function consistently for all $s \in \mathbb{C}$ as a meromorphic function with polar part $\frac{(-1)^n}{n!}\frac{1}{s+n}$ at $s = -n$, $n \in \mathbb{Z}_{\geq 0}$.

The integral (1) is a special case of the **Mellin transform**. Suppose that $\phi(t)$ ($t > 0$) is any function which decays rapidly at infinity (i.e., $\phi(t) = \mathrm{O}(t^{-A})$ as $t \to \infty$ for every $A \in \mathbb{R}$) and blows up at most polynomially at the origin (i.e., $\phi(t) = \mathrm{O}(t^{-C})$ as $t \to 0$ for some $C \in \mathbb{R}$). Then the integral

$$\mathbf{M}\phi(s) = \int_0^\infty \phi(t)\, t^{s-1}\,dt$$

converges absolutely and locally uniformly in the half-plane $\Re(s) > C$ and hence defines a holomorphic function of s in that region. The most frequent situation occurring in number theory is that $\phi(t) = \sum_{n=1}^{\infty} c_n e^{-nt}$ for some complex numbers $\{c_n\}_{n\geq 1}$ which grow at most polynomially in n. Such a function automatically satisfies the growth conditions just specified, and using formula (1) (with t replaced by nt in the integral), we easily find that the Mellin transform $\mathbf{M}\phi(s)$ equals $\Gamma(s)D(s)$, where $D(s) = \sum_{n=1}^{\infty} c_n n^{-s}$ is the Dirichlet series associated to ϕ. Thus the Mellin transformation allows one to pass between Dirichlet series, which are of number-theoretical interest, and exponential series, which are analytically much easier to handle.

Another useful principle is the following. Suppose that our function $\phi(t)$, still supposed to be small as $t \to \infty$, satisfies the functional equation

$$\phi(\frac{1}{t}) = \sum_{j=1}^{J} A_j t^{\lambda_j} + t^h \phi(t) \qquad (t > 0), \tag{2}$$

where h, A_j and λ_j are complex numbers. Then, breaking up the integral defining $\mathbf{M}\phi(s)$ as $\int_0^1 + \int_1^\infty$ and replacing t by t^{-1} in the first term, we find for $\Re(s)$ sufficiently large

$$\begin{aligned}\mathbf{M}\phi(s) &= \int_1^\infty \Big(\sum_{j=1}^{J} A_j t^{\lambda_j} + t^h\phi(t)\Big)\, t^{-s-1}\,dt + \int_1^\infty \phi(t)\, t^{s-1}\,dt \\ &= \sum_{j=1}^{J} \frac{A_j}{s-\lambda_j} + \int_1^\infty \phi(t)\left(t^s + t^{h-s}\right)\frac{dt}{t}.\end{aligned}$$

The second term is convergent for all s and is invariant under $s \mapsto h - s$. The first term is also invariant, since applying the functional equation (2) twice shows that for each j there is a j' with $\lambda_{j'} = h - \lambda_j$, $A_{j'} = -A_j$. Hence we have the

Functional Equation Principle. *If $\phi(t)$ $(t > 0)$ is small at infinity and satisfies the functional equation (2) for some complex numbers h, A_j and λ_j, then the Mellin transform $\mathbf{M}\phi(s)$ has a meromorphic extension to all s and is holomorphic everywhere except for simple poles of residue A_j at $s = \lambda_j$ $(j = 1, \ldots, J)$, and $\mathbf{M}\phi(h - s) = \mathbf{M}\phi(s)$.*

This principle is used to establish most of the functional equations occurring in number theory, the first application being the proof of the functional equation of $\zeta(s)$ given by Riemann in 1859 (take $\phi(t) = \sum_{n=1}^{\infty} e^{-\pi n^2 t}$, so that $\mathbf{M}\phi(s) = \pi^{-s}\Gamma(s)\zeta(2s)$ by what was said above and (2) holds with $h = \frac{1}{2}$, $J = 2$, $\lambda_1 = 0$, $\lambda_2 = \frac{1}{2}$, $A_2 = -A_1 = \frac{1}{2}$ by formula (2) of Appendix A1.

As a final application of the Mellin transform, we prove the formula for $k_s(t)$ stated in Appendix A1. As we just saw, the function λ^{-s} $(\lambda > 0)$ can be written as $\Gamma(s)^{-1}$ times the Mellin transform of $e^{-\lambda t}$. Hence for $a \in \mathbb{R}$ we have $k_s(a) = \Gamma(s)^{-1}\,\mathbf{M}\phi_a(s)$ where

$$\phi_a(t) = \int_{-\infty}^{\infty} e^{-iax}\, e^{-(x^2+1)t}\,dx = \sqrt{\frac{\pi}{t}}\, e^{-t-a^2/4t}.$$

Hence $\pi^{-\frac{1}{2}}\Gamma(s)k_s(a) = \int_0^\infty e^{-t-a^2/4t}\, t^{s-\frac{3}{2}}\, dt$. For $a = 0$ this equals $\Gamma(s-\frac{1}{2})$, while for $a > 0$ it equals $2\left(\frac{a}{2}\right)^{s-\frac{1}{2}} \int_0^\infty e^{-a\cosh u} \cosh(s-\frac{1}{2})u\, du$, as one sees by substituting $t = \frac{1}{2}ae^u$.

A3. Dirichlet characters. If N is a natural number, then a **Dirichlet character modulo** N is a complex valued function χ on $\mathbb{Z}$ satisfying

a) $\chi(n)$ depends only on n modulo N;

b) $\chi(n) = 0$ if and only if $(n, N) > 1$;

c) $\chi(n_1 n_2) = \chi(n_1)\chi(n_2)$ for all $n_1,\, n_2 \in \mathbb{Z}$.

It is easily seen that under these conditions $\chi(1) = 1$ and $\chi(n)$ is a root of unity for any n prime to N. It also follows that $\chi(-1)$ is either $+1$ or -1; in the former case χ is an even function of n (i.e. $\chi(-n) = \chi(n)$ for all n) and is called an **even character**, and in the latter case it is an odd function and is called an **odd character**. The set of Dirichlet characters modulo N forms a group under multiplication which can be identified in the obvious way with the set of group homomorphisms from $(\mathbb{Z}/N\mathbb{Z})^*$ (the multiplicative group of residue classes modulo N which are prime to N) to $\mathbb{C}^*$ (the multiplicative group of non-zero complex numbers).

If N_1 is a divisor of N and χ_1 a Dirichlet character modulo N_1, then the function $\chi : N \to \mathbb{Z}$ defined by

$$\chi(n) = \begin{cases} \chi_1(n) & \text{if } (n, N) = 1 \\ 0 & \text{if } (n, N) > 1 \end{cases}$$

is a Dirichlet character modulo N called the character **induced** by χ_1. A character modulo N which is not induced from one modulo a proper divisor of N is called **primitive**. Any Dirichlet χ modulo N is induced from a unique primitive character modulo a unique divisor N_1 of N, called the **conductor** of χ. As an example, the **principal** (or **trivial**) **character** χ_0 modulo N, defined by $\chi_0(n) = 1$ if $(n, N) = 1$ and $\chi_0(n) = 0$ otherwise, has the conductor 1.

The most important characters are the **quadratic** ones, i.e., those taking on only the values 0, 1 and -1. If $p > 2$ is prime, then the **Legendre symbol** modulo p is the symbol $\left(\frac{\cdot}{p}\right)$ defined by

$$\left(\frac{n}{p}\right) = \begin{cases} +1 & \text{if } n \equiv x^2 \pmod{p} \text{ for some } x \not\equiv 0 \pmod{p}\,, \\ 0 & \text{if } p|n\,, \\ -1 & \text{if } n \text{ is not congruent to a square } \pmod{p}\,; \end{cases}$$

it is easily seen to be a Dirichlet character modulo p. More generally, if $N = p_1^{r_1} \dots p_k^{r_k}$ is a positive odd number, then the **Jacobi symbol** $\left(\frac{\cdot}{N}\right)$, defined by $\left(\frac{n}{N}\right) = \left(\frac{n}{p_1}\right)^{r_1} \dots \left(\frac{n}{p_k}\right)^{r_k}$, is a Dirichlet character modulo N, even or odd according as N is congruent to 1 or 3 modulo 4. Finally, if $D \neq 0$ is a **discriminant** (= integer congruent to 0 or 1 modulo 4), then we have the associated **Kronecker symbol** $\left(\frac{D}{\cdot}\right)$ which is defined by taking $\left(\frac{D}{0}\right) = 0$, $\left(\frac{D}{-1}\right) = \operatorname{sgn}(D)$, $\left(\frac{D}{p}\right)$ equal to the Legendre symbol if p is an odd prime, $\left(\frac{D}{2}\right) = 1$, -1 or 0 according as D is congruent to $1 \pmod{8}$, $5 \pmod{8}$ or $0 \pmod{4}$, and $\left(\frac{D}{n_1 n_2}\right) = \left(\frac{D}{n_1}\right)\left(\frac{D}{n_2}\right)$

for all $n_1,\ n_2 \in \mathbb{Z}$. The new definition of $\left(\frac{D}{n}\right)$ agrees with the Jacobi symbol if n is positive and odd, so there is no ambiguity. The Kronecker symbol is a Dirichlet character modulo $|D|$, the properties (b) and (c) being obvious and the property (a) a consequence of the law of quadratic reciprocity (Exercise 4). It is primitive if and only if D is a **fundamental discriminant** (= square-free integer congruent to 1 modulo 4 or 4 times a square-free integer which is not congruent to 1 modulo 4). In general, we can write D uniquely as $c^2 D_0$ with c a natural number and D_0 a fundamental discriminant, and then $\left(\frac{D}{\cdot}\right)$ is induced from $\left(\frac{D_0}{\cdot}\right)$ and has the conductor D_0. Any quadratic character is induced from the Kronecker symbol $\left(\frac{D}{\cdot}\right)$ for a unique fundamental discriminant D dividing N. There is a 1:1 correspondence between non-trivial primitive quadratic characters and quadratic fields given by assigning to $\left(\frac{D}{\cdot}\right)$ with D fundamental the field $\mathbb{Q}(\sqrt{D})$ and to a quadratic field K of discriminant D the Kronecker symbol $\left(\frac{D}{\cdot}\right)$. This relation can be described more intrinsically in the language of algebraic number theory by saying that χ is the totally multiplicative function whose value at -1 is $+1$ or -1 according as K is a real or imaginary quadratic field and whose value at a prime p is $+1$, -1 or 0 according as p splits into a product of two distinct prime ideals, remains prime, or is the square of a prime ideal in K.

If χ is a Dirichlet character modulo N, then the Dirichlet series $L(s,\chi) = \sum_{n=1}^{\infty} \chi(n)\, n^{-s}$ is called the ***L*-series associated to** χ. This series converges absolutely only for $\Re(s) > 1$, but is known to extend to an analytic function in the entire complex plane (except for a simple pole at $s = 1$ if χ is the principal character), the values at non-negative integers being rational linear combinations of the values $\chi(1), \ldots, \chi(N)$ (and in particular rational numbers if χ is quadratic). Moreover, $L(s,\chi)$ satisfies a **functional equation** which for χ primitive takes the form $L^*(s,\chi) = L^*(1-s,\chi)$ if χ is quadratic and $L^*(s,\chi) = W_\chi L^*(1-s,\bar{\chi})$ in general, where $L^*(s,\chi)$ is defined as $(N/\pi)^{s/2}\Gamma(s/2)L(s,\chi)$ if χ is an even character and as $(N/\pi)^{s/2}\Gamma((s+1)/2)L(s,\chi)$ if χ is odd, $\bar{\chi}$ is the complex conjugate character of χ, and W_χ is a certain complex number of absolute value 1.

Exercises

1. Show that there are exactly $\phi(N)$ Dirichlet characters modulo N, where $\phi(N) = |(\mathbb{Z}/N\mathbb{Z})^*|$ is the Euler totient function. More precisely, show that the group of Dirichlet characters modulo N is isomorphic to the group $(\mathbb{Z}/N\mathbb{Z})^*$ (hint: use the structure theorem for finite abelian groups).

2. Determine the number of *primitive* Dirichlet characters modulo N (hint: show that this number is a multiplicative function of N and compute it for prime powers).

3. Given a Dirichlet character χ modulo N, what is the relation between N, the period of χ as a function on $\mathbb{Z}$, and the conductor of χ? Give examples where these three numbers are (i) all equal, (ii) all distinct.

4. The law of quadratic reciprocity says that for odd primes p and q the Legendre symbols $(\frac{q}{p})$ and $(\frac{p}{q})$ are equal if p or q is congruent to 1 modulo 4 and opposite if both p and q are congruent to 3 modulo 4, while the two supplementary reciprocity laws say that

$\left(\frac{-1}{p}\right)$ is $+1$ for $p \equiv 1 \pmod 4$ and -1 for $p \equiv -1 \pmod 4$ and that $\left(\frac{2}{p}\right)$ is $+1$ for $p \equiv \pm 1 \pmod 8$ and -1 for $p \equiv \pm 3 \pmod 8$. Use these theorems to deduce that the Kronecker symbol $\left(\frac{D}{\cdot}\right)$ is a periodic function of period $|D|$.