Chapter 4

Introduction to Modular Forms

by Don Zagier

Table of Contents

1. A Supply of Modular Forms

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. (For properties of curves of genus 1, see the lectures of Cohen and Bost/Cartier in this volume.) 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 $PSL_2(\mathbb{Z})$ 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** $j(\tau) = e^{-2\pi i\tau} + 744 + 196884e^{2\pi i\tau} + \cdots$ which will be defined below (cf. Section **B**). 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**

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

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. It is a Hilbert space with respect to the **Petersson scalar product**

$$(2) \qquad (f,g) = \iint_{\mathfrak{H}/\Gamma_1} v^k f(\tau)\overline{g(\tau)}\,d\mu \qquad (f,g \in S_k),$$

where we have written τ as $u+iv$ and $d\mu$ for the $SL(2,\mathbb{R})$-invariant measure $v^{-2}\,du\,dv$ on $\mathfrak{H}$.

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. In these notes we will give many illustrations of this type of phenomenon and of the way in which modular forms are used. But to do this we first need to have a supply of modular forms on hand to work with. In this first part a number of constructions of modular forms will be given, the general theory being developed at the same time in the context of these examples.

A Eisenstein series

The first construction 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

$$(3) \qquad G_k(\tau) = \frac{(k-1)!}{2(2\pi i)^k} \sum_{m,n}{}' \frac{1}{(m\tau+n)^k},$$

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 **A**. 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

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

where $\zeta(s)=\sum\limits_{n=1}^{\infty}\frac{1}{n^s}$ is Riemann's zeta function. The number $\frac{(-1)^{k/2}(k-1)!}{(2\pi)^k}\zeta(k)$ is rational and in fact equals $-\frac{B_k}{2k}$, where B_k denotes the kth Bernoulli number (= coefficient of $\frac{x^k}{k!}$ in $\frac{x}{e^x-1}$); 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 (for all of this, cf. the lectures of Bost and Cartier). Putting this into the formula for G_k and collecting for each n the terms with $rm=n$, we find finally

$$G_k(\tau) = -\frac{B_k}{2k} + \sum_{n=1}^{\infty}\sigma_{k-1}(n)q^n = \frac{1}{2}\zeta(1-k) + \sum_{n=1}^{\infty}\sigma_{k-1}(n)q^n, \tag{4}$$

where $\sigma_{k-1}(n)$ denotes $\sum_{r|n} r^{k-1}$ (sum over all positive divisors r of n) and we have used the abbreviation $q=e^{2\pi i\tau}$, **a convention that will be used from now on.**

The right-hand side of (4) 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, since $f(\tau)\,d\tau$ 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). However, its transformation properties under the modular group can be easily determined using **Hecke's trick**: Define a function G_2^* by

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

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 easily to the Fourier development $G_2^*(\tau) = G_2(\tau) + (8\pi v)^{-1}$ ($\tau = u + iv$ as before). 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

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

The beginnings of the Fourier developments of the first few G_k are given by

$$G_2(\tau) = -\frac{1}{24} + q + 3q^2 + 4q^3 + 7q^4 + 6q^5 + 12q^6 + 8q^7 + 15q^8 + \cdots$$
$$G_4(\tau) = \frac{1}{240} + q + 9q^2 + 28q^3 + 73q^4 + 126q^5 + 252q^6 + \cdots$$
$$G_6(\tau) = -\frac{1}{504} + q + 33q^2 + 244q^3 + 1057q^4 + \cdots$$
$$G_8(\tau) = \frac{1}{480} + q + 129q^2 + 2188q^3 + \cdots$$
$$G_{10}(\tau) = -\frac{1}{264} + q + 513q^2 + \cdots$$
$$G_{12}(\tau) = \frac{691}{65520} + q + 2049q^2 + \cdots$$
$$G_{14}(\tau) = -\frac{1}{24} + q + 8193q^2 + \cdots$$

Note that *the Fourier coefficients appearing are all rational numbers*, 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.

B The discriminant function

Define a function Δ in $\mathfrak{H}$ by

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

Then

$$\frac{\Delta'(\tau)}{\Delta(\tau)} = \frac{d}{d\tau}\left(2\pi i\tau + 24\sum_{r=1}^{\infty}\log(1-q^r)\right)$$
$$= 2\pi i\left(1 - 24\sum_{r=1}^{\infty}\frac{rq^r}{1-q^r}\right)$$
$$= -48\pi i\left(-\frac{1}{24} + \sum_{n=1}^{\infty}\Big(\sum_{r|n} r\Big)q^n\right) = -48\pi i G_2(\tau).$$

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

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

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

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

in which only positive exponents of q occur. Hence Δ is a cusp form of weight 12.

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 is empty for k odd (take $a = d = -1$, $b = c = 0$ in (1)), as is M_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, stressed in the introduction to this part, 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; comparing the constant terms gives the proportionality constant as 120 and hence the far from obvious identity

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

Similarly, $(240G_4)^3 - (504G_6)^2$ and Δ are both cusp forms of weight 12 and hence must be proportional. (Cf. Cohen's lectures for the interpretation of this identity in terms of elliptic curves.) In fact, one deduces easily from what has just been said that every modular form is (uniquely) expressible as a polynomial in G_4 and G_6.

Comparing the Fourier expansions of the first few G_k as given in the last section 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

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

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.

C Theta series

We will be fairly brief on this topic, despite its great importance and interest for physicists, because it is treated in more detail in the lectures of Bost and Cartier. 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**

$$\Theta_Q(\tau) = \sum_{x\in\mathbb{Z}^r} q^{Q(x)}.$$

In general this will not be a modular form on the full modular group $\Gamma_1 = PSL_2(\mathbb{Z})$, but on a subgroup of finite index. 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

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

for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(4)$. To prove this, one uses the Poisson summation formula to prove that $\Theta_Q(-1/4\tau) = -2i\tau\Theta_Q(\tau)$; together with the trivial invariance property $\Theta_Q(\tau+1) = \Theta_Q(\tau)$, this shows that Θ_Q is a modular form of weight 1 with respect to the group generated by $\left(\begin{smallmatrix} 0 & -\frac{1}{2} \\ 2 & 0 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, which contains $\Gamma_0(4)$ as a subgroup of index 2.

More generally, if $Q : \mathbb{Z}^r \to \mathbb{Z}$ is any positive definite integer-valued quadratic form in r variables, r even, then Θ_Q is a modular form of weight $r/2$ on some group $\Gamma_0(N) = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1 \mid c \equiv 0 \pmod{N}\}$ with some character $\chi \pmod{N}$, i.e.

$$\Theta_Q\left(\frac{a\tau + b}{c\tau + d}\right) = \chi(d)(c\tau + d)^{r/2}\Theta_Q(\tau) \qquad \text{for all } \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N).$$

The integer N, called the **level** of Q, is determined as follows: write $Q(x) = \frac{1}{2}x^t Ax$ where A is an even symmetric $r\times r$ matrix (i.e., $A = (a_{ij})$, $a_{ij} = a_{ji} \in \mathbb{Z}$, $a_{ii} \in 2\mathbb{Z}$); then N is the smallest positive integer such that NA^{-1} is again even. The character χ is given by $\chi(d) = \left(\frac{D}{d}\right)$ (Kronecker symbol) with $D = (-1)^{r/2} \det A$. For the form $Q(x_1, x_2) = x_1^2 + x_2^2$ above, we have $A = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right)$,

$N = 4$, $\chi(d) = (-1)^{(d-1)/2}$. As a further example, the two quadratic forms $Q_1(x_1, x_2) = x_1^2 + x_1x_2 + 6x_2^2$ and $Q_2(x_1, x_2) = 2x_1^2 + x_1x_2 + 3x_2^2$ have level $N = 23$ and character $\chi(d) = \left(\frac{-23}{d}\right) = \left(\frac{d}{23}\right)$; the sum $\Theta_{Q_1}(\tau) + 2\Theta_{Q_2}(\tau)$ is an Eisenstein series $3+2\sum_{n=1}^{\infty}\left(\sum_{d|n}\chi(d)\right)q^n$ of weight 1 and level 23 (this is a special case of Gauss's theorem on the total number of representations of a natural number by all positive definite binary quadratic forms of a given discriminant), and the difference $\Theta_{Q_1}-\Theta_{Q_2}$ is two times the cusp form $q\prod_{n=1}^{\infty}(1-q^n)(1-q^{23n})$, the 24th root of $\Delta(\tau)\Delta(23\tau)$.

If we want modular forms on the full modular group $\Gamma_1 = PSL_2(\mathbb{Z})$, then we must have $N = 1$ as the level of Q; equivalently, the even symmetric matrix A must be unimodular. This can happen only if the dimension r is divisible by 8 (for a proof using modular forms, cf. Section **D** of Part 2). In dimension 8 there is only one such quadratic form Q up to isomorphism (i.e., up to change of base in $\mathbb{Z}^8$), and Θ_Q is a multiple of the Eisenstein series G_4. In dimension 16 there are two equivalence classes of forms Q, in dimension 24 there are 24, and in dimension 32 there are known to be more than 80 million classes. A theorem of Siegel tells us that the sum of the theta-series attached to all the Q of a given dimension r, each one weighted by a factor $1/|\mathrm{Aut}(Q)|$, is a certain multiple of the Eisenstein series $G_{r/2}$. Notice the applicability of the principle emphasized in the introduction that the finite-dimensionality of the spaces M_k, combined with the existence of modular forms arising from arithmetic situations, gives easy proofs of non-obvious arithmetic facts. For instance, the theta-series of the unique form Q of dimension 8 must be $240G_4$ (since it has weight 4 and starts with 1), so that there are exactly $240\sigma_3(n)$ vectors $x \in \mathbb{Z}^8$ with $Q(x) = n$ for each $n \in \mathbb{N}$; and the two forms of dimension 16 must have the same theta-series (since $\dim M_8 = 1$ and both series start with 1), so they have the same number ($= 480\sigma_7(n)$) of vectors of length n for every n. This latter fact, as noticed by J. Milnor, gives examples of non-isometric manifolds with the same spectrum for the Laplace operator: just take the tori $\mathbb{R}^{16}/\mathbb{Z}^{16}$ with the flat metrics induced by the two quadratic forms in question.

Finally, we can generalize theta series by including **spherical functions**. If $Q : \mathbb{Z}^r \to \mathbb{Z}$ is our quadratic form, then a homogeneous polynomial $P(x) = P(x_1, \ldots, x_r)$ is called 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) for which $Q = \sum y_j^2$, or $\Delta_Q = 2(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_r})A^{-1}(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_r})^t$ in the original coordinate system, 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 space of cusp forms of weight 5 and character $\left(\frac{-4}{\cdot}\right)$ on $\Gamma_0(4)$, and since this space is 1-dimensional it must be of the form

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

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 $(\zeta^t Ax)^\nu$, where $\zeta \in \mathbb{C}^r$ is isotropic (i.e., $Q(\zeta) = \frac{1}{2}\zeta^t A\zeta = 0$). 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. As an example of a spherical theta series of a more general kind we mention Freeman Dyson's identity

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

for the discriminant function Δ of Section **B**.

D Eisenstein series of half-integral weight

In the last section, 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 $f(\frac{a\tau+b}{c\tau+d}) = (\epsilon_{c,d}(c\tau+d)^{\frac{1}{2}})^{2r+1}f(\tau)$ 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 $\sum_{n\ge 0} a(n)q^n$ 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

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

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 $PSL_2(\mathbb{Z})$-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 $PSL_2(\mathbb{Z})$ (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' (cf. Section **A**) 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:

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

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.

E New forms from old

The words 'new' and 'old' here are not being used in their technical sense—introduced in Part 2—but simply to refer to the various methods available for manufacturing modular forms out of previously constructed ones.

The first and 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 many times, as when we compared G_4^2 and G_8. We also found the structure of the graded ring $M_* = \bigoplus M_k$ of all modular forms on the full modular group Γ_1: it is the free $\mathbb{C}$-algebra on two generators G_4 and G_6 of weights 4 and 6. The modular forms on a subgroup $\Gamma \subset \Gamma_1$ of finite index also form a ring. For instance, for $\Gamma = \Gamma_0(2)$ this ring is the free $\mathbb{C}$-algebra on two generators $G_2^{(2)}$ and G_4 of weights 2 and 4, where

$$G_2^{(2)}(\tau) = G_2(\tau) - 2G_2(2\tau) = \frac{1}{24} + \sum_{n=1}^{\infty}\Bigg(\sum_{\substack{d|n\\ d\text{ odd}}} d\Bigg) q^n = \frac{1}{24} + q + q^2 + 4q^3 + \cdots$$

(this is a modular form because $G_2(\tau) - 2G_2(2\tau)$ can also be written as $G_2^*(\tau) - 2G_2^*(2\tau)$, and G_2^* transforms like a modular form of weight 2 on Γ_1). In general, the graded ring of modular forms on Γ will not be a free algebra, but must be given by more than 2 generators and a certain number of relations; it will be free exactly when the Riemann surface $\mathfrak{H}/\Gamma \cup \{\text{cusps}\}$ has genus 0. We also note that the ring of modular forms on Γ contains $M_*(\Gamma_1) = \mathbb{C}[G_4, G_6]$ as a subring and hence can be considered as a module over this ring. As such, it is always *free on n generators*, where n is the index of Γ in Γ_1. For instance, every modular form of weight k on $\Gamma_0(2)$ can be uniquely written as $A(\tau)G_2^{(2)}(\tau) + B(\tau)G_4(\tau) + C(\tau)G_4(2\tau)$ where $A \in M_{k-2}$, $B, C \in M_{k-4}$ (example: $G_2^{(2)}(\tau)^2 = \frac{1}{12}G_4(\tau) + \frac{1}{3}G_4(2\tau)$).

The next method is to apply to two known modular forms f and g of weights k and l **H. Cohen's differential operator**

$$(8)\qquad F_\nu(f,g) = (2\pi i)^{-\nu} \sum_{\mu=0}^{\nu} (-1)^\mu \binom{k+\nu-1}{\mu}\binom{l+\nu-1}{\nu-\mu} f^{(\nu-\mu)} g^{(\mu)},$$

where ν is a nonnegative integer and $f^{(\mu)}$, $g^{(\mu)}$ denote the μth derivatives of f and g. As we will see in a moment, this is a modular form of weight $k+l+2\nu$ on the same group as f and g. For $\nu = 0$ we have $F_0(f,g) = fg$, so the new method is a generalization of the previous one. For $\nu = 1$ we have $F_1(f,g) = \frac{1}{2\pi i}[lf'g - kfg']$; this operation is antisymmetric in f and g and satisfies the Jacobi identity, so that it makes $M_{*-2} = \bigoplus_n M_{n-2}$ into a graded Lie algebra. For ν positive, $F_\nu(f,g)$ has no constant term, so that F_ν maps $M_k \bigotimes M_l$ to $S_{k+l+2\nu}$. The first non-trivial example is $F_1(G_4, G_6) = -\frac{1}{35}\Delta$, 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. (Notice that this identity involves only integers; in general, it is clear that F_ν maps functions with integral or rational Fourier coefficients to another such function.) As another example, observe that applying F_ν to two theta series Θ_{Q_j} associated to quadratic forms $Q_j : \mathbb{Z}^{r_j} \to \mathbb{Z}$ $(j = 1, 2)$ gives rise to a theta-series attached to the form $Q_1 \bigoplus Q_2 : \mathbb{Z}^{r_1+r_2} \to \mathbb{Z}$ and a spherical polynomial of degree ν. For instance, if $\theta(\tau) = \sum q^{n^2}$ is the basic theta-series of weight $\frac{1}{2}$ on $\Gamma_0(4)$, then one checks easily that $\frac{8}{3}F_2(\theta,\theta)$ is the function $\Theta_{Q,P} = \sum_{x_1,x_2\in\mathbb{Z}}(x_1^4 - 6x_1^2x_2^2 + x_2^4)q^{x_1^2+x_2^2}$ discussed at the end of Section **C**. Thus the construction of modular forms via theta-series with spherical functions is a special case of the use of the differential operator F_ν.

We now sketch the proof that F_ν maps modular forms to modular forms. If f is a modular form of weight k on some group Γ, then for $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ and $\mu \in \mathbb{Z}_{\geq 0}$ the formula

$$(9) \qquad f^{(\mu)}\Big(\frac{a\tau + b}{c\tau + d}\Big) = \sum_{\lambda=0}^{\mu} \frac{\mu!(k+\mu-1)!}{\lambda!(\mu-\lambda)!(k+\lambda-1)!} c^{\mu-\lambda}(c\tau+d)^{k+\mu+\lambda} f^{(\lambda)}(\tau)$$

is easily proved by induction on μ (to get from μ to $\mu+1$, just differentiate and multiply by $(c\tau+d)^2$). These transformation formulas can be combined into the single statement that the generating function

$$(10) \qquad \tilde{f}(\tau, X) = \sum_{\mu=0}^{\infty} \frac{1}{\mu!(k+\mu-1)!} f^{(\mu)}(\tau) X^\mu \qquad (\tau \in \mathfrak{H}, X \in \mathbb{C})$$

satisfies

$$(11) \quad \tilde{f}\Big(\frac{a\tau+b}{c\tau+d}, \frac{X}{(c\tau+d)^2}\Big) = (c\tau+d)^k\, e^{cX/(c\tau+d)}\, \tilde{f}(\tau, X) \qquad \Big(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma\Big).$$

Writing down the same formula for a second modular form g of weight l, we find 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.

Finally, we can get new modular forms from old ones by applying the **'slash operator'**

$$f(\tau) \mapsto (f|_k\gamma)(\tau) = (\det\gamma)^{k/2}(c\tau+d)^{-k} f\Big(\frac{a\tau+b}{c\tau+d}\Big)$$

to an f of weight k on Γ, where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is a 2×2 integral matrix which does *not* belong to Γ (if $\gamma \in \Gamma$, of course, then $f|_k\gamma = f$ by definition). This will in general be a modular form on some subgroup of Γ of finite index, but often by combining suitable combinations of images $f|_k\gamma$ we can obtain functions that transform like modular forms on Γ or even on a larger group. Important special cases are the operators

$$V_m f(\tau) = m^{-\frac{k}{2}}\Big(f|_k\begin{pmatrix} m & 0 \\ 0 & 1 \end{pmatrix}\Big) = f(m\tau), \qquad U_m f(\tau) = m^{\frac{k}{2}-1}\sum_{j=1}^{m}\Big(f|_k\begin{pmatrix} 1 & j \\ 0 & m \end{pmatrix}\Big)(\tau)$$

($m \in \mathbb{N}$), which map $\sum a(n)q^n$ to $\sum a(n)q^{mn}$ and to $\sum a(mn)q^n$, respectively. Both map forms of weight k on $\Gamma_0(N)$ to forms of the same weight on $\Gamma_0(mN)$; if m divides N, then U_m even maps forms on $\Gamma_0(N)$ to forms on $\Gamma_0(N)$. Sometimes, applying U_m can even *reduce* the level, which is always a good thing. For instance, if $f = \sum a(n)q^n$ is a modular form of even weight k on $\Gamma_0(4)$, then $U_2 f = \sum a(2n)q^n$ is a modular form of weight k on $\Gamma_0(2)$, and if f has the additional property that $a(n) = 0$ whenever $n \equiv 2 \pmod 4$, then $U_4 f = \sum a(4n)q^n$ even belongs to $M_k = M_k(\Gamma_1)$. Such f occur, for instance, when one multiplies (or applies the operator F_ν to) two forms $g_1 \in M_{r_1+\frac{1}{2}}$, $g_2 \in M_{r_2+\frac{1}{2}}$ with $r_1 + r_2 = k-1$ (resp. $r_1 + r_2 = k - 2\nu - 1$), since then r_1 and r_2 have opposite parity and consequently one of the g's contains only powers q^n with $n \equiv 0$ or $1 \pmod 4$, the other only powers with $n \equiv 0$ or $3 \pmod 4$. This situation will arise in Part 3 in the derivation of the Eichler-Selberg trace formula.

Important operators which can be built up out of the V_m and U_m are the **Hecke operators**, which are the subject of the next part.

F Other sources of modular forms

We have described the main *analytic* ways to produce modular forms on Γ_1 and its subgroups. Another method comes from algebraic geometry: 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 $\Gamma_0(N)$ with $p + 1 - a(p)$ equal to the number of points of the elliptic curve over $\mathbb{F}_p$ for every prime number p. However, this cannot really be considered 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.

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 analytic, algebraic geometric, and number theoretic approaches come together. This is for the special class of modular forms called 'CM' (complex multiplication) forms: 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 $\sum_{x_1,x_2\in\mathbb{Z}}(x_1^4-6x_1^2x_2^2+x_2^4)q^{x_1^2+x_2^2} = q\prod_{n=1}^{\infty}(1-q^n)^{2+2(n,4)}$ which occurred in Section **C**. The characteristic property of these CM forms is that they have highly lacunary Fourier developments. This is because binary quadratic forms represent only a thin subset of all integers (at most $\mathrm{O}\big(x/(\log x)^{1/2}\big)$ integers $\le x$).

Finally, modular forms in one variable can be obtained by restricting in various ways different kinds of modular forms in more than one variable (Jacobi, Hilbert, Siegel, ...), these in turn being constructed by one of the methods of this part. The Jacobi forms will be discussed in Part 4.

2. Hecke Theory

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 **A** and describe their eigenforms and the associated Dirichlet series in Sections **B** and **C**, respectively. The final section of this part describes the modifications of the theory for modular forms on subgroups of $PSL_2(\mathbb{Z})$.

A Hecke operators

At the beginning of Part 1 we introduced the notion of modular forms of higher weight by giving an isomorphism

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

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

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

(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

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

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 $\Gamma_1 = PSL_2(\mathbb{Z})$. Clearly this definition depends on k and we should more correctly write $T_k(n)f$ or (the standard notation) $f|_k T_n$, but we will consider the weight as fixed and write simply $T_n f$ for convenience. In terms of the slash operator

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

introduced in Part 1**E**, formula (3) can be expressed in the form

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

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 $T_n f$ 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 ($(T_n f)|_k\gamma = T_n f$ 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 $\sum_{m=0}^{\infty} a_m q^m$ ($q = e^{2\pi i\tau}$), then the Fourier expansion of $T_n f$ is given by*

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

where $\sum_{d|n,m}$ denotes a sum over the positive common divisors of n and m. In particular, $T_n f$ 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

$$(5) \qquad T_n T_m = \sum_{d|n,m} d^{k-1}\, T_{nm/d^2}.$$

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 $c \neq 0$, then we can choose a matrix $\gamma = \big(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\big) \in PSL_2(\mathbb{Z})$ with $\frac{a'}{c'} = \frac{a}{c}$, 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 $\gamma\big(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\big)$ ($\gamma \in PSL_2(\mathbb{Z})$) 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 $\gamma\binom{a\ b}{0\ d} = \pm\binom{a\ b+dr}{0\ \ d}$, so the choice of μ is unique if we require $a,\ d > 0$ and $0 \le b < d$. Hence

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

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 Part **1A**), 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

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

i.e., $T_p = U_p + p^{k-1} V_p$ where U_p and V_p are the operators defined in **1E**. (More generally, (4) says that T_n for any n is a linear combination of products $U_d V_a$ with $ad = n$.) 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

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

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

B Eigenforms

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

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

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 of Part 1**A**, 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

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

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

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

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, $a(p_1^{r_1} \dots p_l^{r_l}) = a(p_1^{r_1}) \dots a(p_l^{r_l})$ for distinct primes $p_1, \dots, p_l$, 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

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

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 $G_k = -\dfrac{B_k}{2k} + \sum_{m=1}^{\infty} \sigma_{k-1}(m)\, q^m \in M_k$ is a Hecke form for all $k \geq 4$ 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 $1 + p^{k-1} + \dots + p^{\nu(k-1)}$, 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 of Part 1 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 Part 1:

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

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 **A**. 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 Hilbert space structure on S_k introduced in the introduction of Part 1 (eq. (2)). 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 Part 1, Section **B**). 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 $\leq d$. 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 Part **3D**.

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

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

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 $\mathbb{Q}(\sqrt{144169})$.

C L-series

The natural reflex of a number-theorist confronted with a multiplicative function $n \mapsto a(n)$ is to form the Dirichlet series $\sum_{n=1}^{\infty} a(n)\, n^{-s}$, 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 $\prod_{p \text{ prime}} \bigl(\sum_{r \ge 0} a(p^r)\, p^{-rs}\bigr)$. 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

$$L(f,s) = \sum_{m=1}^{\infty} \frac{a(m)}{m^s} \tag{12}$$

(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

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

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

$$a(p^r) = 1 + p^{k-1} + \cdots + p^{r(k-1)} = \frac{p^{(r+1)(k-1)} - 1}{p^{k-1} - 1},$$

$$A_p(x) = \sum_{r=0}^{\infty} \frac{p^{(r+1)(k-1)} - 1}{p^{k-1} - 1} x^r = \frac{1}{\big(1 - p^{k-1}x\big)\big(1 - x\big)}$$

$$L(G_k, s) = \prod_p \frac{1}{1 - \sigma_{k-1}(p)p^{-s} + p^{k-1-2s}} = \prod_p \frac{1}{(1 - p^{k-1-s})(1 - p^{-s})}$$
$$= \zeta(s - k + 1)\zeta(s),$$

where $\zeta(s)$ is the Riemann zeta function. (Of course, we could see this directly: the coefficient of n^{-s} in $\zeta(s - k + 1)\zeta(s) = \sum\limits_{d,e\geq 1} \dfrac{d^{k-1}}{(de)^s}$ 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*

$$\text{(13)} \quad a(n) = O\big(n^{k-1}\big) \qquad (f \in M_k), \qquad\qquad a(n) = O\big(n^{\frac{k}{2}}\big) \qquad (f \in S_k).$$

Hence the L-series $L(f, s)$ converges absolutely and locally uniformly in the half-plane $\Re(s) > k$ in any case and in the larger half-plane $\Re(s) > \frac{k}{2} + 1$ 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 $\sigma_{k-1}(n) = n^{k-1}\sum_{d|n} d^{-k+1} < n^{k-1}\sum_{d=1}^{\infty} d^{-k+1} < 2n^{k-1}$ because $k > 2$), 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 (13). 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 part, is due to Hecke.)

(ii) This follows immediately from the 'functional equation principle' in Appendix **B**, since the function

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

is exponentially small at infinity and satisfies the functional equation

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

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 (13) is clearly the 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 explained in the next part. 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

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

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; 5 years later he proved the Weil conjectures, thus establishing (14). Using the form of the generating function $A_p(x)$ given above, one sees that (14) is equivalent to

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

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

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

The proof of (16) 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.

D Forms of higher level

In most of these notes, we restrict attention to the full modular group $\Gamma_1 = PSL_2(\mathbb{Z})$ 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 restrict attention to the subgroups $\Gamma_0(N) = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1 \mid c \equiv 0 \pmod{N}\}$ introduced in Part 1.

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 $T_p = U_p$ and $T_{p^r} = (T_p)^r$ rather than $T_p = U_p + p^{k-1} V_p$ and a 3-term recursion relation for $\{T_{p^r}\}$. For general n, 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

$$\begin{aligned} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) &\Rightarrow \begin{pmatrix} a & bM \\ c/M & d \end{pmatrix} \in \Gamma_0(N') \\ &\Rightarrow f\Bigl(M \frac{a\tau + b}{c\tau + d}\Bigr) = f\Bigl(\frac{a(M\tau) + bM}{(c/M)(M\tau) + d}\Bigr) = (c\tau + d)^k f(M\tau). \end{aligned}$$

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 Part 1**A**, 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 $M_k(\Gamma_0(N))^{\text{old}}$ of old forms and a certainly complementary space $M_k(\Gamma_0(N))^{\text{new}}$ (for cusp forms, $S_k(\Gamma_0(N))^{\text{new}}$ is just the orthogonal complement of $S_k(\Gamma_0(N))^{\text{old}}$ with respect to the Petersson scalar product), and if we define a **Hecke form of level** N to be a form in $M_k(\Gamma_0(N))^{\text{new}}$ 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 $M_k(\Gamma_0(N))^{\text{new}}$, and their Fourier coefficients are real algebraic integers as before. For the pth Fourier coefficient (p prime) of a Hecke form in $S_k(\Gamma_0(N))^{\text{new}}$ 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 $\pm p^{(k-1)/2}$ otherwise. Finally, there is no overlapping between the new forms 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

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

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 level 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, there are in general further splittings. The general situation is that $S_k(\Gamma_0(N))^{\text{new}}$ splits as the sum of subspaces of some dimensions $d_1, \ldots, d_r \geq 1$, 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

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

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 Part 1F) to the elliptic curve $y^2 - y = x^3 - x^2$, i.e., $p - a(p)$ gives the number of solutions of $y^2 - y = x^3 - x^2$ in integers modulo any prime p.

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

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

with coefficients in $\mathbb{Z} + \mathbb{Z}\dfrac{1+\sqrt{5}}{2}$ and the conjugate form

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

obtained by replacing $\sqrt{5}$ by $-\sqrt{5}$ everywhere in f_1.

3. $k = 2$, $N = 37$. Again $\dim M_k(\Gamma_0(N))^{\rm old}$ is spanned by $G_2^*(\tau) - NG_2^*(N\tau)$ and $M_k(\Gamma_0(N))^{\rm new} = S_k(\Gamma_0(N))^{\rm new}$ 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 $\mathbb{Z}$; they correspond 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 $\dim M_k(\Gamma_0(N))^{\rm old}$ is spanned by the two Eisenstein series $G_4(\tau)$ and $G_4(N\tau)$ and $M_k(\Gamma_0(N))^{\rm new} = S_k(\Gamma_0(N))^{\rm new}$ is 3-dimensional, spanned by the forms

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

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 must be modified slightly: it is now

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

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 $v \mapsto \frac{1}{v}$. 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 asymptotically $\frac{1}{2}$ 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.

3. The Rankin-Selberg Method and its Applications

The Rankin-Selberg convolution method is one of the most powerful tools in the theory of automorphic forms. In this part we explain two principal variants of it—one involving non-holomorphic Eisenstein series and one involving only the holomorphic Eisenstein series constructed in Part 1. We will also give several applications, the most important one being a proof of the formula of Eichler and Selberg for the traces of Hecke operators acting on spaces of holomorphic modular forms. 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.

A 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 Part 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 $2^{s+1}d^{-s/2}G(\tau, s)$ 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 **B**) implies

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

On the other hand, the Poisson summation formula (cf. Appendix **A**) implies that $\Theta_\tau(\frac{1}{t}) = t\Theta_\tau(t)$, so the function $\phi(t) = \frac{1}{2}(\Theta_\tau(t) - 1)$ satisfies $\phi(t^{-1}) = -\frac{1}{2} + \frac{1}{2}t + t\phi(\frac{1}{t})$. The 'functional equation principle' formulated in Appendix **B** 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 Part 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

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

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

$$\begin{aligned} G(\tau, s) = \zeta(2s)v^s + \frac{\pi^{\frac{1}{2}}\Gamma(s - \frac{1}{2})}{\Gamma(s)} v^{1-s} \sum_{m=1}^{\infty} m^{1-2s} \\ + \frac{2\pi^s}{\Gamma(s)} v^{\frac{1}{2}} \sum_{\substack{m \geq 1 \\ r \neq 0}} m^{\frac{1}{2}-s} |r|^{s-\frac{1}{2}} K_{s-\frac{1}{2}}(2\pi m|r|v)\, e^{2\pi i m r u}, \end{aligned}$$

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

$$G^*(\tau, s) = \zeta^*(2s)v^s + \zeta^*(2s - 1)v^{1-s} + 2v^{\frac{1}{2}} \sum_{n \neq 0} \sigma^*_{s-\frac{1}{2}}(|n|) K_{s-\frac{1}{2}}(2\pi|n|v) e^{2\pi i n u},$$

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 **B**), $\sigma^*_{-\nu}(n) = \sigma^*_\nu(n)$, and $K_{-\nu}(t) = K_\nu(t)$. □

As an immediate consequence of the Fourier development of G^* and the identity $K_{\frac{1}{2}}(t) = \sqrt{\pi/2t}\, e^{-t}$, we find

$$\begin{aligned} \lim_{s\to 1}\left(G^*(\tau, s) - \frac{1/2}{s - 1} \right) &= \frac{\pi}{6}v - \frac{1}{2}\log v + C + 2 \sum_{m,r=1}^{\infty} \frac{1}{m} \Re(e^{2\pi i m r \tau}) \\ &= \frac{\pi}{6}v - \frac{1}{2}\log v + C - \sum_{r=1}^{\infty} \log|1 - e^{2\pi i r\tau}|^2 \\ &= -\frac{1}{24}\log\left(v^{12}|\Delta(\tau)|^2\right) + C, \end{aligned}$$

where $C = \lim_{s\to 1}(\zeta^*(s) - (s-1)^{-1})$ is a certain constant (in fact given by $\frac{1}{2}\gamma - \frac{1}{2}\log 4\pi$, where γ is Euler's constant) and $\Delta(\tau)$ the discriminant function of Part 1. This formula is called the **Kronecker limit formula** and has many applications in number theory. Together with the invariance of $G(\tau, s)$ under $PSL_2(\mathbb{Z})$, it leads to another proof of the modular transformation property of $\Delta(\tau)$.

B 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 **B**) of the constant term $c_0(v)$ of F:

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

(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

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

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 Section **A**:

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

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 $\{\pm\left(\begin{smallmatrix}1&n\\0&1\end{smallmatrix}\right) \mid n \in \mathbb{Z}\}$ 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

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

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 $\int_0^\infty\left(\int_0^1 F(u+iv)\,du\right)v^{s-2}\,dv$, 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 Part **1B**, $k = 12$). 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

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

If f is a Hecke form, then the coefficients $a(n)$ real and $\sum a(n)^2 n^{-s-k+1} = \zeta(s) \sum a(n^2) n^{-s-k+1}$ 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

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

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 $G\big(\tau, \frac{1}{2}(1+i\alpha)\big)$. 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 $s = \frac{1}{2}(1 - i\alpha)$,

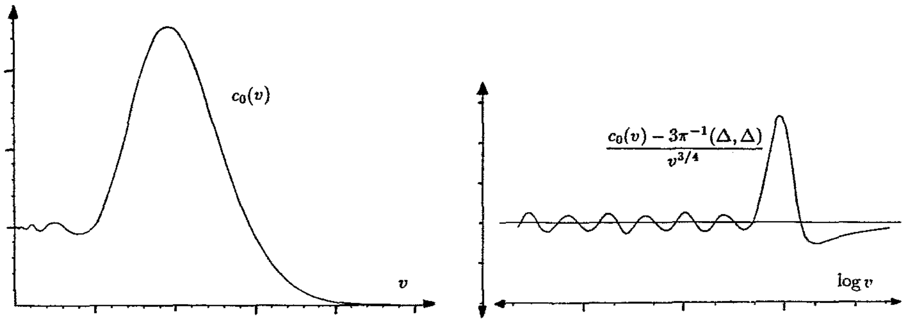

Fig. 1. The constant term $c_0(v) = v^{12}\sum_{n=1}^{\infty}\tau(n)^2e^{-4\pi nv}$.

$G(\tau, s) = \overline{F(\tau)}$, 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 $c_0(v) = C_0v^{\lambda_0} + C_1v^{\lambda_1} + C_2v^{\lambda_2} + \cdots$ 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 $v^{12}\sum\tau(n)^2e^{-4\pi nv}$ 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 $\zeta(s)$. 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 $O(v^{\frac{1}{4}+\epsilon})$ if the Riemann hypothesis is true and by $O(v^{\frac{1}{2}+\epsilon})$ 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.

C The Rankin-Selberg method (holomorphic case)

The calculations here are very similar to those of Section **B**, 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

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

and consequently

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

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 **E** of Part 1, where now $h+l+2\nu=k$. Here we find

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

where in the last line we have used formula (9) of Part 1, **E**. The same argument as before now leads to

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

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.

D Application: The Eichler-Selberg trace formula

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 Part 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

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

Theorem. (Eichler, Selberg) *Let $H(N)$ $(N \geq 0)$ be the Kronecker-Hurwitz class numbers defined in* **D** *of Part 1 and denote by $p_k(t,n)$ the homogeneous polynomial*

$$p_k(t,n) = \sum_{0 \leq r \leq \frac{k}{2}-1} \binom{k-2-r}{r} (-n)^r t^{k-2-2r} = Coeff_{X^{k-2}}\left(\frac{1}{1-tX+nX^2}\right)$$

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 (for $k > 2$) follows from (1), since

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

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

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

where the terms '…' 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 $H(N)$.

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 $(a_i(n_j))_{1\le i,j\le d}$ 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. The formula for $\mathrm{Tr}(T(n))$ thus gives an algorithm for obtaining all cusp forms of a given weight (and level).

We now sketch a proof of the Eichler-Selberg trace formula. The basic tool we will use is the 'holomorphic version' of the Rankin-Selberg method proved in the last section, 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 Section **C** 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 $PSL_2(\mathbb{Z})$. They can be simplified by using the operator U_4 introduced in Section **E** of Part 1 and replacing $F_\nu(G_h, g)$ by $U_4(F_\nu(G_h, g))$, which belongs to $M_k(\Gamma_1)$ if $g \in M_l$, $h, l \in \mathbb{Z} + \frac{1}{2}$, $k = h + l + 2\nu \in 2\mathbb{Z}$ (cf. comments at the end of Part 1, **E**). In this situation, formula (2) of Section **C** 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)
$$(f, U_4(F_\nu(G_{r+\frac{1}{2}}, \theta))) = c_{\nu,r} \sum_{n=1}^{\infty} \frac{a(n^2)}{(n^2)^{k-\nu-1}} \quad (r > 1 \text{ odd}, \ \nu \ge 0, \ k = r+2\nu+1),$$

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 $G^*_{3/2}$ which was defined in Part 1**D**. The function $U_4(F_\nu(G^*_{3/2}, \theta))$ is no longer holomorphic, but we can apply the 'holomorphic projection operator' (cf. Appendix **C**) 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 **B**, 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)
$$\pi_{\mathrm{hol}}(U_4(F_\nu(G^*_{3/2}, \theta))) = c_k\, \mathbf{t}(\tau) + c'_k G_k(\tau)$$

for some constant c'_k.

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

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

and hence

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

so

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

(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 $v^{k-2} e^{-4\pi n v}$ and integrated from $v = 0$ to $v = \infty$. The integral turns out to be elementary and one finds after a little calculation

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

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.

4. Jacobi forms

When we introduced modular forms, we started with functions F of lattices $\Lambda \subset \mathbb{C}$ invariant under rescaling $\Lambda \mapsto \lambda\Lambda$ ($\lambda \in \mathbb{C}^*$); these corresponded via $f(\tau) = F(\mathbb{Z}\tau + \mathbb{Z})$ to modular functions. The quotient $\mathbb{C}/\Lambda$ is an elliptic curve, so we can think of F (or f) as functions of elliptic curves. It is natural to make them functions *on* elliptic curves as well, i.e., to consider functions Φ which depend both on Λ and on a variable $z \in \mathbb{C}/\Lambda$. The equations

$$\Phi(\lambda\Lambda, \lambda z) = \Phi(\Lambda, z), \quad \Phi(\Lambda, z+\omega) = \Phi(\Lambda, z) \qquad (\lambda \in \mathbb{C}^\times,\ \omega \in \Lambda)$$

correspond via $\phi(\tau, z) = \Phi(\mathbb{Z}\tau + \mathbb{Z}, z)$ to functions ϕ on $\mathfrak{H} \times \mathbb{C}$ satisfying

$$(1) \qquad \begin{aligned} &\phi\Big(\frac{a\tau+b}{c\tau+d}, \frac{z}{c\tau+d}\Big) = \phi(\tau, z), \\ &\phi(\tau, z+\ell\tau+m) = \phi(\tau, z) \qquad \Big(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_1,\ \ell, m \in \mathbb{Z}\Big). \end{aligned}$$

We call a meromorphic function ϕ on $\mathfrak{H} \times \mathbb{C}$ satisfying (1) a **Jacobi function.**

However, there can clearly never be a holomorphic Jacobi function, since by Liouville's theorem a holomorphic function on $\mathbb{C}$ invariant under all transformations $z \mapsto z + \omega$ ($\omega \in \Lambda$) must be constant. Thus, just as the concept of modular function was too restrictive and had to be extended to the concept of modular forms of weight k, corresponding to functions on lattices transforming under $\Lambda \mapsto \lambda\Lambda$ with a scaling factor λ^{-k}, the concept of Jacobi functions must be extended by incorporating appropriate scaling factors into the definition. The right requirements, motivated by examples which will be presented in Section **A**, turn out to be

$$(2) \qquad \phi\Big(\frac{a\tau+b}{c\tau+d}, \frac{z}{c\tau+d}\Big) = (c\tau+d)^k\, e^{\frac{2\pi i N c z^2}{c\tau+d}}\, \phi(\tau, z) \qquad \Big(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_1\Big)$$

and

$$(3) \qquad \phi(\tau, z+\ell\tau+m) = e^{-2\pi i N(\ell^2\tau + 2\ell z)}\phi(\tau, z) \qquad (\ell, m \in \mathbb{Z}),$$

where N is a certain integer. Finally, just as with modular forms, there must be a growth condition at infinity; it turns out that the right condition here is to require that ϕ have a Fourier expansion of the form

$$(4) \qquad \phi(\tau, z) = \sum_{n=0}^{\infty} \sum_{\substack{r \in \mathbb{Z} \\ r^2 \le 4Nn}} c(n, r)\, q^n\, \zeta^r \qquad (q = e^{2\pi i\tau},\ \zeta = e^{2\pi i z})$$

(again, the rather odd-looking condition $r^2 \le 4Nn$ will be motivated by the examples). A function $\phi : \mathfrak{H} \to \mathbb{C}$ satisfying the conditions (2), (3) and (4) will be called a **Jacobi form** of **weight** k and **index** N.

Surprisingly, in most of the occurrences of modular forms and functions in physics—in particular, those connected with theta functions and with Kac-Moody algebras—it is actually Jacobi forms and functions which are involved. It is for this reason, and because the theory is not widely known, that we have devoted an entire part to these functions.

A Examples of Jacobi forms

The simplest theta series, namely the function

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

introduced at the beginning of Section 1**D**, is actually just the specialization to $z = 0$ ('Thetanullwert') of the two-variable function

$$\theta(\tau, z) = \sum_{n\in\mathbb{Z}} q^{n^2}\zeta^{2n} = 1 + (\zeta^2+\zeta^{-2})\,q + (\zeta^4+\zeta^{-4})\,q^4 + (\zeta^6+\zeta^{-6})\,q^9 + \cdots,$$

and similarly the transformation equation

$$\theta\Bigl(\frac{a\tau+b}{c\tau+d}\Bigr) = \epsilon_{c,d}\,(c\tau+d)^{\frac12}\theta(\tau) \qquad \Bigl(\begin{pmatrix} a & b\\ c & d\end{pmatrix} \in \Gamma_0(4),\ \epsilon_{c,d}^4 = 1\Bigr)$$

is just the specialization to $z = 0$ of the more general transformation equation

$$\theta\Bigl(\frac{a\tau+b}{c\tau+d}, \frac{z}{c\tau+d}\Bigr) = \epsilon_{c,d}\,(c\tau+d)^{\frac12}\, e^{\frac{2\pi i c z^2}{c\tau+d}}\,\theta(\tau, z) \qquad \Bigl(\begin{pmatrix} a & b\\ c & d\end{pmatrix} \in \Gamma_0(4)\Bigr).$$

It is also easily checked that θ satisfies

$$\theta(\tau, z + \ell\tau + m) = e^{-2\pi i(\ell^2\tau + 2\ell z)}\,\theta(\tau, z)$$

(just replace n by $n+\ell$ in the summation defining θ), so that $\theta(\tau, z)$ is, with the obvious modifications in the definition given before, a Jacobi form of weight 1/2 and index 1 on the group $\Gamma_0(4)$. The function $\theta(\tau, z)$ is one of the classical Jacobi theta functions and this is the reason for the name 'Jacobi form.'

Just as for the one-variable theta functions discussed in Part 1, if we want to get forms of integral weight and on the full modular group, rather than of weight 1/2 and on $\Gamma_0(4)$, we must start with quadratic forms in an even number of variables and whose associated matrix has determinant 1. If $Q : \mathbb{Z}^{2k} \to \mathbb{Z}$ is a positive definite quadratic form in $2k$ variables given by an even symmetric unimodular matrix A (i.e. $Q(x) = \frac12 x^t A x$, $a_{ij} \in \mathbb{Z}$, $\frac12 a_{ii} \in \mathbb{Z}$, $\det A = 1$), then for each vector $y \in \mathbb{Z}^{2k}$ the theta-function

$$\Theta_{Q,y}(\tau, z) = \sum_{x\in\mathbb{Z}^{2k}} q^{Q(x)}\,\zeta^{B(x,y)}, \tag{5}$$

where $\zeta = e^{2\pi i z}$ as before and $B(x, y) = x^t A y$ is the bilinear form associated to Q, is a Jacobi form of weight k and index $N = Q(y)$. The transformation law (2) is proved using the Poisson summation formula as for the special case $\Theta_Q(\tau) = \Theta_{Q,0}(\tau, 0)$ studied in Part 1; the transformation law (3) is proved directly from the expansion (5) by making the substitution $x \mapsto x + \ell y$; and the form of the Fourier expansion required in (4) is clear from (5) and the Cauchy-Schwarz inequality $B(x,y)^2 \le 4Q(x)Q(y)$. (This motivates the inequality $r^2 \le 4Nn$ in (4), as promised.)

The next example is that of Eisenstein series. The Eisenstein series of Part 1 can be written as

$$G_k(\tau) = \frac{1}{2}\zeta(1-k) \sum_{\gamma \in \Gamma_\infty \backslash \Gamma_1} 1|_k \gamma(\tau),$$

where $|_k$ is the slash operator introduced in **1E** and the summation is over the cosets of $\Gamma_\infty = \left\{\pm\left(\begin{smallmatrix}1 & n\\ 0 & 1\end{smallmatrix}\right) \mid n \in \mathbb{Z}\right\}$ in $\Gamma_1 \subset PSL_2(\mathbb{Z})$ (cf. Part 3). In the Jacobi form context we must generalize the slash operator to a new operator $|_{k,N}$ defined by

$$\big(\phi|_{k,N}\gamma\big)(\tau,z) = (c\tau+d)^{-k}\, e^{-\frac{2\pi i N c z^2}{c\tau+d}}\, \phi\Big(\frac{a\tau+b}{c\tau+d}, \frac{z}{c\tau+d}\Big) \qquad \Big(\gamma = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) \in \Gamma_1\Big),$$

$$\big(\phi|_{k,N}[\ell,m]\big)(\tau,z) = e^{2\pi i N(\ell^2\tau + 2\ell z)}\, \phi(\tau, z+\ell\tau+m) \qquad (\ell, m \in \mathbb{Z})$$

(so that $\phi|_{k,N}\gamma = \phi|_{k,N}[\ell,m] = \phi$ if ϕ is a Jacobi form of weight k and index N). We then define an Eisenstein series

$$G_{k,N}(\tau,z) = \zeta(3-2k) \sum_{\gamma\in\Gamma_\infty\backslash\Gamma_1} \sum_{\ell\in\mathbb{Z}} \big((1|_{k,N}\gamma)|_{k,N}[\ell,0]\big)(\tau,z)$$

or more explicitly

$$G_{k,N}(\tau,z) = \frac{1}{2}\zeta(3-2k) \sum_{\substack{c,d,\ell\in\mathbb{Z}\\ (c,d)=1}} \frac{e^{2\pi i N\left(\ell^2\gamma_{c,d}(\tau) + \frac{2\ell z}{c\tau+d} - \frac{cz^2}{c\tau+d}\right)}}{(c\tau+d)^k},$$

where $\gamma_{c,d}$ for each pair of coprime integers c, d denotes an element of $PSL_2(\mathbb{Z})$ with lower row $(c\ d)$. The series is convergent for $k > 2$ and defines a Jacobi form of weight k and index N. Moreover, its Fourier expansion can be computed by a calculation analogous to, though somewhat harder than, the one given in **1A**. The result is that the Fourier coefficients are rational numbers of arithmetic interest, expressible in closed form in terms of the function $H(r,n)$ introduced in Section **D** of Part 1 in connection with Eisenstein series of half-integral weight. In particular, for $N=1$ the result is simply

$$G_{k,1}(\tau,z) = \sum_{n=0}^{\infty} \sum_{|r|\le\sqrt{4n}} H(k-1, 4n-r^2)\, e^{2\pi i(n\tau+rz)}.$$

That the coefficient of $q^n\zeta^r$ depends only on $4n-r^2$ is not an accident: it is easily seen that the transformation equation (3) in the case $N=1$ is equivalent to the condition that the Fourier coefficient $c(n,r)$ as defined in (4) depend only on $4n-r^2$, while for general N (3) is equivalent to the requirement that $c(n,r)$ depend only on $4n-r^2$ and on the residue of r modulo $2N$. The fact that the coefficients of the Jacobi Eisenstein series were essentially the same as the coefficients of Eisenstein series in one variable but of half-integral weight is also not accidental: There is in fact an intimate connection between Jacobi forms and modular forms of half-integral weight, obtained by associating to the Jacobi form ϕ the collection of functions $\sum_d C_\mu(d)\,q^d$, $\mu=1,2,\ldots,2N$, where $C_\mu(d)$ is the common value of the $c(n,r)$ with $4n-r^2=d$ and $r\equiv\mu \pmod{2N}$; each of these $2N$ functions is a modular form of weight $k-\frac{1}{2}$ with respect to some subgroup of Γ_1, and the entire $2N$-tuple satisfies a transformation law with respect to the whole group Γ_1. However, we do not elaborate on this here.

The beginnings of the Fourier expansions of the first few Jacobi Eisenstein series (of index 1) are

$$\begin{aligned} G_{4,1}(\tau,z) &= -\tfrac{1}{252} + \left(-\tfrac{1}{252}\zeta^2 - \tfrac{2}{9}\zeta - \tfrac{1}{2} - \tfrac{2}{9}\zeta^{-1} - \tfrac{1}{252}\zeta^{-2}\right)q \\ &\quad + \left(-\tfrac{1}{2}\zeta^2 - \tfrac{16}{7}\zeta - 3 - \tfrac{16}{7}\zeta^{-1} - \tfrac{1}{2}\zeta^{-2}\right)q^2 \\ &\quad + \left(-\tfrac{2}{9}\zeta^3 - 3\zeta^2 - 6\zeta - \tfrac{74}{9} - 6\zeta^{-1} - 3\zeta^{-2} - \tfrac{2}{9}\zeta^{-3}\right)q^3 + \cdots, \\ G_{6,1}(\tau,z) &= -\tfrac{1}{132} + \left(-\tfrac{1}{132}\zeta^2 + \tfrac{2}{3}\zeta + \tfrac{5}{2} + \tfrac{2}{3}\zeta^{-1} - \tfrac{1}{132}\zeta^{-2}\right)q \\ &\quad + \left(\tfrac{5}{2}\zeta^2 + 32\zeta + 57 + 32\zeta^{-1} + \tfrac{5}{2}\zeta^{-2}\right)q^2 + \cdots, \\ G_{8,1}(\tau,z) &= -\tfrac{1}{12} + \left(-\tfrac{1}{12}\zeta^2 - \tfrac{14}{3}\zeta - \tfrac{61}{2} - \tfrac{14}{3}\zeta^{-1} - \tfrac{1}{12}\zeta^{-2}\right)q + \cdots. \end{aligned}$$

To get more examples, we can combine these in various ways. In particular, the two functions

$$\begin{aligned} \phi_{10,1}(\tau,z) &= 882\,G_6(\tau)\,G_{4,1}(\tau,z) + 220\,G_4(\tau)\,G_{6,1}(\tau,z), \\ \phi_{12,1}(\tau,z) &= -840\,G_8(\tau)\,G_{4,1}(\tau,z) - 462\,G_6(\tau)\,G_{6,1}(\tau,z) \end{aligned} \tag{6}$$

are Jacobi forms of index 1 and weights 10 and 12, respectively, and in fact are Jacobi *cusp* forms (i.e. $n>0$, $r^2<4Nn$ in (4)) with Fourier expansions starting

$$\begin{aligned} \phi_{10,1}(\tau,z) &= \left(\zeta - 2 + \zeta^{-1}\right)q + \left(-2\zeta^2 - 16\zeta + 36 - 16\zeta^{-1} - 2\zeta^{-2}\right)q^2 \;\cdots, \\ \phi_{12,1}(\tau,z) &= \left(\zeta + 10 + \zeta^{-1}\right)q + \left(10\zeta^2 - 88\zeta - 132 - 88\zeta^{-1} + 10\zeta^{-2}\right)q^2 \;\cdots; \end{aligned}$$

their ratio $\phi_{12,1}/\phi_{10,1}$ is $-3\pi^{-2}$ times the **Weierstrass $\wp$-function** $\wp(z;\mathbb{Z}\tau+\mathbb{Z})$ from the theory of elliptic functions (cf. the lectures of Cohen and Bost/Cartier in this volume).

Other important examples of Jacobi forms are obtained from the Fourier developments of Siegel modular forms on the symplectic group $Sp(2,\mathbb{Z})$, but we cannot go into this here since we have not developed the theory of Siegel modular forms.

B Known results

In this section we describe a few highlights from the theory of Jacobi forms.

(i) If $\phi \not\equiv 0$ is a Jacobi form of weight k and index N, then it is easily seen by integrating $\frac{d}{dz}\log\phi$ around a fundamental parallelogram for $\mathbb{C}/(\mathbb{Z}\tau+\mathbb{Z})$ that ϕ has exactly $2N$ zeros in this parallelogram (here we are considering τ as fixed and ϕ as a function of z alone). In particular, ϕ cannot have a zero of multiplicity greater than $2N$ at the origin, so in the Taylor expansion

$$\phi(\tau,z)=\chi_0(\tau)+\chi_1(\tau)z+\chi_2(\tau)z^2+\cdots$$

the first $2N+1$ coefficients determine ϕ completely. On the other hand, one easily sees by differentiating (2) repeatedly with respect to z and then setting z equal to 0 that χ_0 is a modular form in τ of weight k (this, of course, is obvious), χ_1 a modular form of weight $k+1$, $\chi_2-\frac{2\pi iN}{k}\chi_0'$ a modular form of weight $k+2$, and more generally

$$\xi_\nu(\tau)=\sum_{0\le\mu\le\nu/2}\frac{(-2\pi iN)^\mu\,(k+\nu-\mu-2)!}{(k+\nu-2)!\,\mu!}\,\chi_{\nu-2\mu}^{(\mu)}(\tau) \tag{7}$$

a modular form of weight $k+\nu$ for every integer $\nu\ge0$. The fact that ϕ is determined by its first $2N+1$ Taylor coefficients means that we have an injective map from the space $J_{k,N}$ of Jacobi forms of weight k and index N into the direct sum $M_k\oplus M_{k+2}\oplus\cdots\oplus M_{k+2N}$ if k is even or $M_{k+1}\oplus M_{k+3}\oplus\cdots M_{k+2N-1}$ if k is odd. In particular, $J_{k,N}$ is finite dimensional, of dimension at most $\frac{1}{12}kN+\mathrm{O}(N^2)$.

The function ξ_ν defined by (7) has the Fourier expansion

$$\xi_\nu(\tau)=(2\pi i)^\nu\,\frac{(k-2)!}{(k+\nu-2)!}\sum_{n=1}^{\infty}\Bigg(\sum_{|r|\le\sqrt{4Nn}}p_{k-1,\nu}(r,Nn)\,c(n,r)\Bigg)\,q^n\,,$$

where the $c(n,r)$ are the coefficients defined by (4) and $p_{d,\nu}(a,b)$ denotes the coefficient of X^ν in $(1-aX+bX^2)^{-d}$. The fact that ξ_ν is a modular form is related to the heat equation operator $8\pi iN\frac{\partial}{\partial\tau}-\frac{\partial^2}{\partial z^2}$, and also to the formula (11) of 1E.

(ii) The bigraded ring of all Jacobi forms (of all weights and indexes) is not finitely generated, since the forms obtained as polynomials in any finite collection would have a bounded ratio of k to N and there is an Eisenstein series $G_{k,1}$ for all $k>2$. However, if we enlarge the space $J_{k,N}$ to the space $\tilde{J}_{k,N}$ of 'weak Jacobi forms,' defined as functions $\phi:\mathcal{H}\times\mathbb{C}\to\mathbb{C}$ satisfying the properties (2)–(4) but with the condition '$r^2\le4Nn$' dropped in (4), then the bigraded ring $\bigoplus_{k,N}\tilde{J}_{k,N}$ is simply the ring of all polynomials in the four functions $G_4(\tau)$, $G_6(\tau)$, $\phi_{10,1}(\tau,z)/\Delta(\tau)$ and $\phi_{12,1}(\tau,z)/\Delta(\tau)$ (with $\phi_{10,1}$ and

$\phi_{12,1}$ as in (6) and Δ as in 1**B**) of weight 4, 6, -2 and 0 and index 0, 0, 1 and 1, respectively. In particular, $\Delta(\tau)^N\phi(\tau,z)$ is a polynomial in $G_4(\tau)$, $G_6(\tau)$, $G_{4,1}(\tau,z)$ and $G_{6,1}(\tau,z)$ for any Jacobi form ϕ of index N.

(iii) There are no Jacobi forms of weight 1 on $PSL_2(\mathbb{Z})$, i.e., $J_{1,N}=\{0\}$ for all N.

(iv) One can define Hecke operators on the spaces $J_{k,N}$ and compute their traces. These turn out to be related to the traces of Hecke operators on the spaces of ordinary modular forms of weight $2k-2$ and level N. Using these, one can construct lifting maps from $J_{k,N}$ to a certain subspace $\mathfrak{M}_{2k-2}(N)\subset M_{2k-2}(\Gamma_0(N))$ which is canonically defined and invariant under all Hecke operators. Moreover, $J_{k,N}$ turns out to be isomorphic to the subspace of forms in $\mathfrak{M}_{2k-2}(N)$ whose Hecke L-series satisfy a functional equation with a minus sign, i.e., to the intersection of $\mathfrak{M}_{2k-2}(N)$ with $M_{2k-2}^{(-1)^k}(\Gamma_0(N))$ (cf. 2**D**).

(v) There is another kind of Jacobi form, called **skew-holomorphic Jacobi forms**, for which statements analogous to those in (iv) hold but with the isomorphism now between the space of skew-holomorphic Jacobi forms and the subspace of forms in $\mathfrak{M}_{2k-2}(N)$ having a *plus* sign in the functional equation of their L-series. By definition, a skew-holomorphic Jacobi form of weight k and index N is a function ϕ on $\mathcal{H}\times\mathbb{C}$ which satisfies the transformation equations (2) and (3) but with $(c\tau+d)^k$ replaced by $(c\bar\tau+d)^{k-1}|c\tau+d|$ in (2) and which has a Fourier expansion like the one in (4) but with the condition $r^2\le 4Nn$ replaced by $r^2\ge 4Nn$ and with $q^n\zeta^r$ multiplied by $e^{-\pi(r^2-4Nn)v/N}$ $(v=\Im(\tau))$. Such a function is again holomorphic in z, but the Cauchy-Riemann condition $\dfrac{\partial\phi}{\partial\bar\tau}=0$ of holomorphy in τ is replaced by the heat equation $\dfrac{\partial\phi}{\partial\tau}=\dfrac{1}{8\pi iN}\dfrac{\partial^2\phi}{\partial z^2}$. The Fourier expansion together with the transformation property (3) can be written uniformly in the holomorphic and non-holomorphic case as

$$(8)\qquad \phi(\tau,z)=\sum_{\substack{r,\Delta\in\mathbb{Z}\\ r^2\equiv\Delta\ (\mathrm{mod}\ 4N)}} C(\Delta,r)\,e^{2\pi i\left(\frac{r^2-\Delta}{4N}\Re(\tau)+i\frac{r^2+|\Delta|}{4N}\Im(\tau)+rz\right)}$$

where $C(\Delta,r)$ depends only on Δ and on $r\pmod{4N}$ and vanishes for $\Delta>0$ (holomorphic case) or $\Delta<0$ (non-holomorphic case).

(vi) There are explicit constructions of Jacobi and skew-Jacobi forms in terms of binary quadratic forms, due to Skoruppa. For instance, if we define $C(\Delta,r)=\sum\mathrm{sgn}(a)$, where the (finite) sum is over all binary quadratic forms $[a,b,c]=ax^2+bxy+cy^2$ of discriminant $b^2-4ac=\Delta$ with $a\equiv 0\pmod N$, $b\equiv r\pmod{4N}$ and $ac<0$, then (8) defines a skew-holomorphic Jacobi form of weight 2 and index N.

Appendices

The following appendices describe some analytic tools useful in the theory of modular forms.

A 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**

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

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 $\varphi(x) = e^{-\pi a x^2}$ with $a > 0$. Then $\hat{\varphi}(t) = a^{-\frac{1}{2}} e^{-\pi t^2/a}$, so we get

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

(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

$$k_s(t) = \begin{cases} \frac{2\pi^{\frac{1}{2}}}{\Gamma(s)} \left(\frac{|t|}{2}\right)^{s-\frac{1}{2}} K_{s-\frac{1}{2}}(|t|) & \text{if } t \neq 0, \\ \frac{\pi^{\frac{1}{2}}\Gamma(s-\frac{1}{2})}{\Gamma(s)} & \text{if } t = 0 \end{cases}$$

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

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

This formula is used for computing the Fourier development of the non-holomorphic Eisenstein series (Part **3A**).

B 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**

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

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., $\phi(t) = O(t^{-A})$ as $t \to \infty$ for every $A \in \mathbb{R}$) and blows up at most polynomially at the origin (i.e., $\phi(t) = O(t^{-C})$ 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

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

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

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

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

Proposition. (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 $s = \lambda_j$ $(j = 1, \dots, J)$, 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 **A**).

As a final application of the Mellin transform, we prove the formula for $k_s(t)$ stated in Appendix **A**. 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 $t = \frac{1}{2}ae^u$.

C 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. Our object is to derive a formula for the Fourier coefficients of $\pi_{\text{hol}}(\Phi)$.

To do this, we introduce the **Poincaré series**. For each integer $m \in \mathbb{N}$ set

$$P_m(\tau) = \sum_{\gamma = \left(\begin{smallmatrix} \cdot & \cdot \\ c & d \end{smallmatrix}\right) \in \Gamma_\infty \backslash \Gamma_1} \frac{e^{2\pi i m \gamma(\tau)}}{(c\tau + d)^k} \qquad \left(\gamma(\tau) = \frac{a\tau + b}{c\tau + d} \text{ for } \gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\right),$$

where the summation is over left cosets of $\Gamma_\infty = \left\{\pm\left(\begin{smallmatrix} 1 & b \\ 0 & 1 \end{smallmatrix}\right), b \in \mathbb{Z}\right\}$ in Γ_1. The series converges absolutely if $k > 2$ and defines a cusp form of weight k. The same unfolding argument as in the Rankin-Selberg method (Part 3, **B**) shows that for a form $f = \sum_1^\infty a(n)q^n \in S_k$ the Petersson scalar product (f, P_m) is given by

$$\begin{aligned}(f,P_m) &= \iint_{\mathfrak{H}/\Gamma_1} f(\tau)\,\overline{P_m(\tau)}\,v^k\,\frac{du\,dv}{v^2}\\ &= \iint_{\mathfrak{H}/\Gamma_\infty} f(\tau)\,\overline{e^{2\pi i m\tau}}\,v^k\,\frac{du\,dv}{v^2}\\ &= \int_0^\infty \left(\int_0^1 f(u+iv)\,e^{-2\pi i m u}\,du\right) e^{-2\pi m v}v^{k-2}\,dv\\ &= \int_0^\infty \left(a(m)e^{-2\pi m v}\right) e^{-2\pi m v}v^{k-2}\,dv\\ &= \frac{(k-2)!}{(4\pi m)^{k-1}}a(m).\end{aligned}$$

In other words, $(4\pi m)^{k-1}P_m(\tau)/(k-2)!$ is the cusp form dual to the operator of taking the mth Fourier coefficient of a holomorphic cusp form.

Now let $\sum_{n\in\mathbb{Z}} c_n(v)e^{2\pi i n u}$ denote the Fourier development of our function $\Phi(\tau)$ and $\sum_{n=1}^\infty c_n q^n$ that of its holomorphic projection to S_k. Then

$$\frac{(k-2)!}{(4\pi m)^{k-1}}c_m = (\pi_{\mathrm{hol}}(\Phi),P_m) = \overline{(P_m,\pi_{\mathrm{hol}}(\Phi))} = \overline{(P_m,\Phi)} = (\Phi,P_m)$$

by the property of P_m just proved and the defining property of $\pi_{\mathrm{hol}}(\Phi)$. Unfolding as before, we find

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

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_m) remains convergent after replacing Φ by its absolute value and P_m by its majorant $\hat{P}_m(\tau)=\sum_{\Gamma_\infty\backslash\Gamma_1}|(c\tau+d)^{-k}e^{2\pi i m\gamma(\tau)}|$. We have

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

with $G(\tau,\frac{k}{2})$ the non-holomorphic Eisenstein series introduced in **A**, Part 3. The estimate there shows that $G(\tau,\frac{k}{2})-\zeta(k)v^{k/2}=\mathrm{O}(v^{1-k/2})$ as $v\to\infty$, so $\hat{P}_m(\tau)=\mathrm{O}(v^{1-k})$. The convergence of $\iint_{\mathfrak{H}/\Gamma_1}|\Phi|\hat{P}_m v^{k-2}\,du\,dv$ 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 $m=0$ (G_k

is proportional to P_0). We have thus proved the following result, first stated by J. Sturm under slightly different hypotheses:

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

(1) *'(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*
(2) *'(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 Part **1A**. 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

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

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 $\sum c_n q^n \in M_4$ 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 $\sum_0^\infty c_n q^n \in M_4 = \langle G_4 \rangle$, we must have $c_n = 240 c_0 \sigma_3(n) = \frac{5}{12}\sigma_3(n)$ for all $n > 0$, an identity that the reader can check for small values of n.

Similarly, if $f = \sum_0^\infty a_n q^n$ is a modular form of weight $l \geq 4$, then $\Phi = f G_2^*$ satisfies the hypotheses of the lemma with $k = l + 2$, $c_0 = -\frac{1}{24} a_0$ and $\epsilon = 1$, and we find that $\pi_{\mathrm{hol}}(f G_2^*) = f G_2 + \frac{1}{4\pi i l} f' + \in M_{l+2}$.

References

We will not attempt to give a complete bibliography, but rather will indicate some places where the interested reader can learn in more detail about the theory of modular forms.

Three short introductions to modular forms can be especially recommended:

(i) the little book *Lectures on Modular Forms* by R.C. Gunning (Princeton, Ann. of Math. Studies **48**, 1962), which in 86 widely spaced pages describes the classical analytic theory and in particular the construction of Poincaré series and theta series,

(ii) Chapter 7 of J-P. Serre's 'Cours d'Arithmétique' (Presses Universitaires de France 1970; English translation: Springer, GTM **7**, 1973), which gives among other things a very clear introduction to the theory of Hecke operators and to the applications of theta series to the arithmetic of quadratic forms, and

(iii) the survey article by A. Ogg in *Modular Functions of One Variable. I* (Springer, Lecture Notes **320**, 1973), which treats some of the modern aspects of the theory and in particular the connection with elliptic curves. (The other volumes in this series, SLN **349**, **350**, **475**, **601**, and **627**, describe many of the developments of the years 1970–76, when the subject experienced a renascence after a long period of dormancy.)

Of the full-length books on the subject, the best introduction is probably Serge Lang's *Introduction to Modular Forms* (Springer, Grundlehren **222**, 1976), which treats both the analytic and the algebraic aspects of theory. It also includes a detailed derivation of the trace formula for Hecke operators on the full modular group (this is in an appendix by me and unfortunately contains an error, corrected in the volume SLN **627** referred to above). Other texts include Ogg's *Modular Forms and Dirichlet Series* (Benjamin 1969), which gives in great detail the correspondence between modular forms and Dirichlet series having appropriate functional equations, as well as an excellent presentation of the theory of theta series with spherical polynomial coefficients, G. Shimura's *Introduction to the Arithmetic Theory of Automorphic Functions* (Princeton 1971), which is more advanced and more heavily arithmetic than the other references discussed here, and the recent book *Modular Forms* by T. Miyake (Springer 1989), which contains a detailed derivation of the trace formula for the standard congruence subgroups of Γ_1. Another good book that treats the connection with elliptic curves and also the theory of modular forms of half-integral weight is N. Koblitz's *Introduction to Elliptic Curves and Modular Forms* (Springer, GTM **97**, 1984). Finally, anyone who really wants to learn the subject from the inside can do no better than to study Hecke's *Mathematische Werke* (Vandenhoeck 1959).

We also mention some books on subjects closely related to the theory of modular forms: for a classically oriented account of the theory of modular functions, Rankin's *Modular Forms and Functions* (Cambridge 1977) or Schoeneberg's *Elliptic Modular Functions: An Introduction* (Springer, Grundlehren **203**, 1974); for the theory of elliptic curves, Silverman's book *The Arithmetic of Elliptic curves* (Springer, GTM **106**, 1986); for the modern point of view on modular forms in terms of the representation theory of $GL(2)$ over the adeles of a number field, Gelbart's *Automorphic Forms on Adele Groups* (Princeton, Ann. of Math. Studies **83**, 1975) or, to go further, *Automorphic Forms, Representations, and L-Functions* (AMS 1979).

We now give in a little more detail sources for the specific subjects treated in these notes.

Part 1. The basic definitions of modular forms and the construction of the Eisenstein series G_k and the discriminant function Δ are given in essentially every introduction. Serre (op. cit.) gives a construction of Δ which is related to, but different from, the one given here: Instead of using the non-holomorphic modular form G_2^*, he uses G_2 itself but analyzes the effect on the value of the non-absolutely convergent series $\sum(m\tau+n)^{-2}$ of summing over m and n in different orders. This approach goes back to Eisenstein. The reader should beware of the fact that Serre normalizes the weight differently, so that, e.g., Δ has weight 6 instead of 12. The best treatment of theta series in the simplest case, namely when the underlying quadratic form is unimodular and there are no spherical coefficients, is also given in Serre's book, but for the general case one must go to Gunning's or (better) Ogg's book, as already mentioned. The Eisenstein series of half-integral weight are already a more specialized topic and are not to be found in any of the books mentioned so far. The construction of the Fourier coefficients of the particular Eisenstein series $G_{r+\frac{1}{2}}$ which we discuss (these are the simplest half-integral-weight series, but there are others) is due to H. Cohen (Math. Ann. **217**, 1975), for $r > 1$, while the construction of the series $G^*_{\frac{3}{2}}$ is contained in an article by Hirzebruch and myself (Inv. math. **36**, 1976, pp. 91–96). The development of the general theory of modular forms of half-integral weight, and in particular the construction of a 'lifting map' from these forms to forms of integral weight, is given in famous papers by G. Shimura (Ann. of Math. **97**, 1973 and in the above-mentioned Lecture Notes **320**); an elementary account of this theory is given in Koblitz's book cited above. Of the constructions described in Section **E**, the differential operator F_ν is constructed in the paper of H. Cohen just cited, but is in fact a special case of more general differential operators constructed by Rankin several years earlier, while the 'slash operators' and the operators V_m and U_m are treated in any discussion of Hecke operators for congruence subgroups of $SL(2,\mathbb{Z})$ and in particular in Chapter VII of Lang's book. Finally, the topics touched upon in Section **F** are discussed in a variety of places in the literature: the connection between modular forms of weight 2 and elliptic curves of weight 2 is discussed e.g. in Silverman's book or the Springer Lecture Notes **476** cited above; the theorem of Deligne and Serre appeared in Ann. Sc. Ec. Norm. Sup. 1974; and the theory of complex multiplication is discussed in Lang's book of the same name and in many other places.

Part 2. As already mentioned, the clearest introduction to Hecke operators for the full modular group is the one in Serre's book, the L-series and their functional equations are the main topic of Ogg's Benjamin book. The theory in the higher level case was first worked out by Atkin and Lehner (Math. Ann. **185**, 1970) and is presented in detail in Chapters VII-VIII of Lang's textbook. Some tables of eigenforms for weight 2 are given in the Lecture Notes volume **476** cited above.

Part 3. The classical reference for the function $G(\tau,s)$ and the Rankin-Selberg method is Rankin's original paper (Proc. Camb. Phil. Soc. **35**, 1939). However, the main emphasis there is on analytic number theory and the derivation of the estimate $a(n) = O(n^{\frac{k}{2}-\frac{1}{5}})$ for the Fourier coefficients $a(n)$ of a cusp form f of weight k (specifically, Δ of weight 12). Expositions of the general method have been given by several authors, including the present one on several occasions (e.g. in two articles in *Automorphic Forms, Representation Theory and Arithmetic*, Springer 1981, and in a paper in J. Fac. Sci. Tokyo **28**, 1982; these also contain the applications mentioned in Section **B**). The proof of the Eichler-Selberg trace formula sketched in Section **D** has not been presented before. Standard proofs can be found in the books of Lang and Miyake, as already mentioned, as well, of course, as in the original papers of Eichler and Selberg.

Part 4. The theory of Jacobi forms was developed systematically in a book by M. Eichler and myself (Progress in Math.**55**, Birkhäuser 1985); special examples, of course, had been known for a long time. The results described in (iii), (v) and (vi) of Section **B** are due to N.-P. Skoruppa (in particular, the construction mentioned in (vi) is to appear in Inv. math. 1990), while the trace formula and lifting maps mentioned under (iv) are joint work of Skoruppa and myself (J. reine angew. Math. **393**, 1989, and Inv. math. **94**, 1988). A survey of these and some other recent developments is given in Skoruppa's paper in the proceedings of the Conference on Automorphic Functions and their Applications, Khabarovsk 1988.

Appendices. The material in Sections **A** and **B** is standard and can be found in many books on analysis or analytic number theory. The method of holomorphic projection was first given explicitly by J. Sturm (Bull. AMS **2**, 1980); his proof is somewhat different from the one we give.