From quadratic functions to modular functions

D. Zagier

Abstract. We study certain functions defined in a very simple way as sums of powers of quadratic polynomials with integer coefficients and discover that these functions have several surprising properties and are related to many other subjects, including Diophantine approximation, special values of zeta functions, modular forms, and Dedekind sums.

Part I. Sums of quadratic polynomials

1. An extremely simple function

We start with the following very elementary construction. Pick a positive integer, say 5. Now for any real number x we consider all quadratic functions with integer coefficients and discriminant 5 which are negative at infinity and positive at x, i.e. $Q(X) = aX^2 + bX + c$ with $a, b, c \in \mathbb{Z}$, $b^2 - 4ac = 5$, $a < 0$, and $ax^2 + bx + c > 0$. By definition the number $Q(x)$ is positive for each such function Q, and we ask for the sum of these numbers, i.e., for the value of the function

$$A(x) = \sum_{\substack{\mathrm{disc}(Q)=5 \\ Q(x)>0>Q(\infty)}} Q(x). \tag{1}$$

We will see below that this sum converges, so that A is a well-defined function $\mathbb{R} \to \mathbb{R}$.

First Surprise. *The function* $A(x)$ *has the constant value* 2.

We illustrate this by calculating a few values of $A(x)$ numerically. In general it is not obvious how to do this, since we do not yet know anything about the convergence of the series. But when x is rational there is no problem, since then the sum in (1) is actually *finite*. For instance, if the function $Q(X) = [a, b, c] := aX^2 + bX + c$ occurs in the sum (1) for $x = 0$ then $a < 0 < c$, so $5 = b^2 + 4|a||c|$ and Q must be either $[-1, 1, 1]$ or $[-1, -1, 1]$; and more generally, if Q occurs for $x = p/q$ then $5q^2 = |bq - 2ap|^2 + 4|a||ap^2 + bpq + cq^2|$, which bounds each of a, b and c. In this way we can find the functions Q occurring in (1) and compute the value of $A(x)$ for any rational value of x. The results for the three simplest values $x = 0$, $\frac{1}{2}$ and $\frac{1}{3}$ are given by the following tables:

Q	$Q(0)$
$[-1,1,1]$	1
$[-1,-1,1]$	1
Sum:	2

Q	$Q(1/2)$
$[-1,1,1]$	5/4
$[-1,-1,1]$	1/4
$[-1,3,-1]$	1/4
$[-5,5,-1]$	1/4
Sum:	2

Q	$Q(1/3)$
$[-1,1,1]$	11/9
$[-1,-1,1]$	5/9
$[-5,5,-1]$	1/9
$[-11,7,-1]$	1/9
Sum:	2

What about irrational values of x? Here the identity $5 = (2ax - b)^2 - 4aQ(x)$ and the inequalities $a < 0 < Q(x)$ imply that $|b - 2ax| < \sqrt{5}$, so there are at most five values of b for any given value of a. We can therefore write a computer program in which a runs over the values $-1, -2, \ldots, -M$ for some large integer M; for each a and each integer $b \in (2ax - \sqrt{5}, 2ax + \sqrt{5})$ the computer checks whether $b^2 \equiv 5 \pmod{4a}$ and, if this is true, prints out the form $Q = [a, b, (b^2 - 5)/4a]$ and the value of $Q(x)$. For $x = 1/\pi$, for instance, a computer search up to the bound $M = 100,000$ found only six functions Q satisfying the conditions $a < 0 < Q(x)$, viz., the four occurring in the above table for 1/3 (which is, after all, very near to $1/\pi$) and the two further functions $[-409, 259, -41]$ and $[-541, 345, -55]$. Adding up the corresponding values of $Q(1/\pi)$ gave (to five places beyond the decimal point) the sum 1.99998. This gave numerical support to the assertion that $A(x) = 2$ for all x and also suggested that the non-occurrence of forms in the range $541 < |a| < 100,000$ was only a temporary effect and that if we went on we would find further quadratic forms whose values at $1/\pi$ add up to the missing 0.00002. Indeed, extending our search to $M = 300,000$ we find two further Q's and the following table:

Q	$Q(1/\pi)$
$[-1,1,1]$	1.21699
$[-1,-1,1]$	0.58037
$[-5,5,-1]$	0.08494
$[-11,7,-1]$	0.11364
$[-409,259,-41]$	0.00190
$[-541,345,-55]$	0.00215
$[-117731,74951,-11929]$	0.00001
$[-133351,84893,-13511]$	0.00001
Sum:	2.00000

Of course, if $A(1/\pi)$ truly is equal to 2, then the sequence of Q's can never terminate, since then $1/\pi$ would satisfy a quadratic equation $Q_1(x)+\ldots+Q_n(x) = 2$ over $\mathbb{Z}$. (This equation cannot be trivial because each Q_i has a negative leading coefficient.) The same argument applies to any other transcendental number, or any algebraic number of degree > 2, so our "Surprise" has a surprising corollary, of which we will indicate a direct proof in §10:

Corollary. *Let x be a real number which is not rational or quadratic over $\mathbb{Q}$. Then there are infinitely many quadratic forms $Q(X) = aX^2 + bX + c$ of discriminant 5 with $a < 0$ and $Q(x) > 0$.*

We close this section by a graph illustrating the terms contributing to the sum (1) for $x = 0$, $1/3$, $1/2$ and 1 and the constancy of the function $A(x)$. The rest of the paper will be devoted to explaining and generalizing this phenomenon.

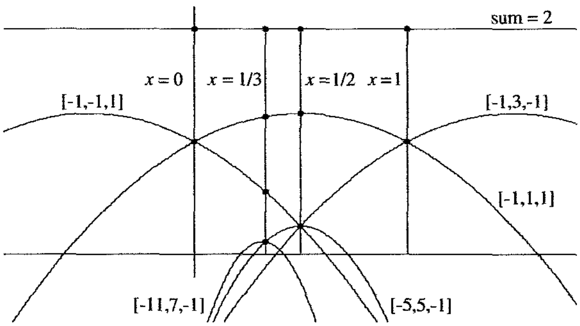

Figure 1

2. Generalization to other discriminants, and special values of L-series

There is, of course, nothing special about the choice of discriminant 5, and we can generalize the function $A(x)$ by defining

$$A_D(x) = \sum_{\substack{a,b,c\in\mathbb{Z},\ a<0,\\ b^2-4ac=D}} \max(0, ax^2 + bx + c) \quad (x \in \mathbb{R}). \tag{2}$$

Here D is a positive integer which is congruent to 0 or 1 modulo 4 (otherwise the sum is empty) and not a perfect square. (The case of square D is somewhat different and will be treated in §7.) The "surprise" of §1 then remains true, except for the value of the constant. We state this as a theorem, the proof of which will be given in §6.

Theorem 1. *For every D the function $A_D(x)$ has a constant value α_D.*

We can verify this experimentally, and at the same time find the value of α_D, by the same algorithm as in §1. (In particular, the same proof as given there shows that the sum defining $A_D(x)$ is finite whenever x is rational.) Here is a table of values up to $D = 30$:

D	5	8	12	13	17	20	21	24	28	29
α_D	2	5	10	10	20	22	20	30	40	30

To get an explicit formula for α_D, we use the theorem to write α_D as $A_D(0)$. Then, recalling that the $[a, b, c]$ occurring in (2) for $x = 0$ are those with $c > 0 > a$, we find the formula

$$\alpha_D = \sum_{b \in \mathbb{Z},\ |b| < \sqrt{D},\ b \equiv D \pmod 2} \sigma_1\left(\frac{D - b^2}{4}\right), \tag{3}$$

where $\sigma_1(n)$ denotes the sum of the divisors of a positive integer n. And suddenly our elementary construction has led us into the realm of more serious mathematics, for the expression occurring in (3) has occurred previously in the literature as a formula for an important invariant of algebraic number theory. To explain this, we suppose that D is a fundamental discriminant, i.e., that it is the discriminant of a real quadratic field K. (The general case will be treated later.) Associated to D are two important Dirichlet series: the Dirichlet L-series $L_D(s) = L(s, \chi_D)$ of the character $\chi_D(n) = (D/n)$ (Kronecker symbol) and the Dedekind zeta-function $\zeta_D(s) = \zeta_K(s)$ of the quadratic field K. They are defined for $\Re(s) > 1$ by

$$\begin{aligned} L_D(s) = L(s, \chi_D) &= \sum_{n=1}^{\infty} \frac{\chi_D(n)}{n^s}, \\ \zeta_D(s) = \zeta_K(s) &= \sum_{\mathfrak{a}} \frac{1}{N(\mathfrak{a})^s} \end{aligned} \tag{4}$$

(where $\mathfrak{a}$ in the second sum ranges over non-zero ideals of $\mathcal{O}_K$) and for other $s \in \mathbb{C}$ by analytic continuation (resp. meromorphic continuation in the case of $\zeta_K(s)$, which has a simple pole at $s = 1$), and satisfy the functional equations

$$\begin{aligned} L_D(1-s) &= \frac{D^{s-1/2}\Gamma(\frac{s}{2})}{\pi^{s-1/2}\Gamma(\frac{1-s}{2})} L_D(s), \\ \zeta_D(1-s) &= \frac{D^{s-1/2}\Gamma(\frac{s}{2})^2}{\pi^{2s-1}\Gamma(\frac{1-s}{2})^2} \zeta_D(s) \end{aligned} \tag{5}$$

and the relation

$$\zeta_D(s) = \zeta(s) L_D(s), \tag{6}$$

where $\zeta(s)$ is the Riemann zeta function. From these we obtain in particular the equalities

$$L_D(-1) = -12\zeta_D(-1) = -\frac{D^{3/2}}{2\pi^2}L_D(2) = -\frac{3D^{3/2}}{\pi^4}\zeta_D(2).$$

It has been known for a long time that the number occurring in this formula is rational and can be given, for instance, by the formula $-\sum_{n=1}^{D-1}\chi_D(n)n^2/2D$. In 1969 Siegel [Si] found a completely different formula based on the study of Eisenstein series for the Hilbert modular group associated to K. His formula (as simplified by Cohen [Co1], [Co2] and myself [Z2]; the original version was somewhat more complicated) expresses $\zeta_K(-1)$ as 1/60 of the expression on the right-hand side of (3). So we have the following

Supplement to Theorem 1. *Suppose that D is the discriminant of a real quadratic field K. Then the number α_D occurring in Theorem 1 is given by*

$$\alpha_D = -5L(-1,\chi_D) = 60\zeta_K(-1). \tag{7}$$

The proof of (7) which we gave here was purely computational, depending on the explicit formula (3) and Siegel's formula for $\zeta_K(-1)$. Later (in §8) we will see the intrinsic reason why α_D is related to $\zeta_K(-1)$.

3. Second generalization: cubes instead of first powers

Being mathematicians, we naturally generalize again. This time, instead of changing the "5" in (1) to another value, we try a different power of the summand, defining

$$B(x) = \sum_{\substack{a,b,c\in\mathbb{Z},\, a<0,\\ b^2-4ac=5}} \max\bigl(0,(ax^2+bx+c)^3\bigr) \quad (x\in\mathbb{R}), \tag{8}$$

Note that the series (8), unlike (1), converges uniformly for all x, since $Q(x) = O(1/a)$ and the number of contributing Q's per value of a is bounded. And now we find:

Second Surprise. *The function $B(x)$ also has the constant value* 2.

We can check this numerically the same way as we did for $A(x)$ in §1, the sum being finite for x rational and convergent (indeed, far more rapidly convergent than before) for x irrational. For instance, cubing the values of $Q(x)$ listed in §1 for the three rational numbers 0, $\frac{1}{2}$ and $\frac{1}{3}$, we find $B(0) = 1 + 1 = 2$, $B(\frac{1}{2}) = \frac{125+1+1+1}{64} = 2$, $B(\frac{1}{3}) = \frac{1331+125+1+1}{729} = 2$, and similarly for $x = 1/\pi$ the values tabulated in §1 give

Q	$Q(1/\pi)^3$
$[-1,1,1]$	1.80243511584123887281298722803
$[-1,-1,1]$	0.19548456121216942259483309670
$[-5,5,-1]$	0.00061290145140734893903713558
$[-11,7,-1]$	0.00146740473128344118486462838
$[-409,259,-41]$	0.00000000682021406025480759889
$[-541,345,-55]$	0.00000000994368585373007613108
$[-117731,74951,-11929]$	0.000000000000000050535245763529
$[-133351,84893,-13511]$	0.000000000000000049513093654604
Sum:	2.000000000000000000000000000000

The convergence is very fast, with just 8 terms of the series (8) giving us 30 digits of accuracy.

Of course, just as in §2, we can generalize to discriminants other than 5, setting

$$B_D(x) = \sum_{\substack{a,b,c\in\mathbb{Z},\ a<0,\\ b^2-4ac=D}} \max(0, (ax^2+bx+c)^3) \quad (x \in \mathbb{R}), \tag{9}$$

and we again find that these functions are always constant, although—contrary to what our "Second surprise" might suggest—the values are not in general the same as those of A_D. In other words, we have:

Theorem 2. *For every D the function $B_D(x)$ has a constant value β_D.*

The number β_D is given by a formula exactly analogous to equation (3), namely:

$$\beta_D = \sum_{b\in\mathbb{Z},\ |b|<\sqrt{D},\ b\equiv D \pmod 2} \sigma_3\Big(\frac{D-b^2}{4}\Big), \tag{10}$$

where $\sigma_3(n)$ for a positive integer n denotes the sum of the cubes of the divisors of n. Again, the results of [Si], [Co2] and [Z2] tell us that this is equal to a special value of a Dirichlet series, namely (for D a fundamental discriminant)

$$\beta_D = L_D(-3) = 120\zeta_D(-3) = \frac{3D^{7/2}}{4\pi^4}L_D(4) = \frac{135D^{7/2}}{2\pi^8}\zeta_D(4). \tag{11}$$

Here is a table of the first few values of the numbers β_D.

D	5	8	12	13	17	20	21	24	28	29
β_D	2	11	46	58	164	274	308	522	904	942

Finally, we should answer a question which the reader might have had: why cubes and not squares? One answer, of course, is that it works. A more illuminating answer is that the function $Q(x)^3$, like $Q(x)$, is positive only on an interval of finite length, whereas $Q(x)^2$ is always ≥ 0, so that the sum with $\max(0, Q(x)^2)$ would always diverge. Of course, if we wrote $\max(0, Q(x))^2$ instead of $\max(0, Q(x)^2)$, then we would get a convergent series, but this is a less natural expression. In fact, it is possible to do something with even powers as well, but only by looking at a more complicated sum. See §9.

4. A new phenomenon

It is now clear how the pattern should go on. We replace the function (1), for instance, by the corresponding sum of fifth powers,

$$C(x) = \sum_{\substack{a,b,c\in\mathbb{Z},\, a<0,\\ b^2-4ac=5}} \max(0, (ax^2+bx+c)^5) \quad (x \in \mathbb{R}), \tag{12}$$

and more generally define $C_D(x)$ by the same sum but with discriminant D instead of discriminant 5. Then $C(x)$ should have the constant value 2, and more generally $C_D(x)$ should be a constant function of x for every D, its value being the integer γ_D defined by

$$\gamma_D = \sum_{b\in\mathbb{Z},\, |b|<\sqrt{D},\, b\equiv D \pmod 2} \sigma_5\left(\frac{D-b^2}{4}\right), \tag{13}$$

where $\sigma_5(n)$ denotes the sum of the 5th powers of the divisors of n.

Unfortunately, this nice theory is not true. Instead, we have:

Third Surprise. *The function $C(x)$ is* not *a constant function of x.*

Of course this is easy to check. For instance, from the values given in §1 we see that $C(0) = 1^5 + 1^5 = 2$ but

$$C\left(\frac{1}{2}\right) = \frac{5^5 + 1^5 + 1^5 + 1^5}{4^5} = \frac{3128}{1024} \neq 2.$$

Indeed, we see without doing any calculation at all that $\sum Q(x)^m$ could not be a constant function of x for every (odd) value of m, because the value of this sum for $x = 0$ equals $1^m + 1^m = 2$, while for other rational values of x we get the sums of the mth powers of other finite collection of rational numbers, and these clearly cannot always be independent of m.

But if the function $C_D(x)$ is not constant, what is it? Since we expected the value of $C_D(x)$ to be equal to the number γ_D defined by (13), it is reasonable to look at the values of the difference $C_D^0(x) := C_D(x) - \gamma_D$. Calculating in the way explained in §1 for a few values of x (we choose rational x so that the sum is finite and the value of $C_D^0(x)$ is also a rational number), we obtain the following table:

D	5	8	12	13
$\gamma_D = C_D(0)$	2	35	310	490
$C_D^0(\frac{1}{2})$	135/128	−135/64	405/32	−1485/128
$C_D^0(\frac{1}{3})$	5120/6561	−10240/6561	20480/2187	−56320/6561
$C_D^0(\frac{1}{4})$	16875/32768	−16875/16384	50625/8192	−185625/32768

An inspection of these values leads to our

Fourth Surprise. *The functions $C_D^0(x)$ for different discriminants D are proportional.*

Stated differently, this says that the function $C_D(x)$ has the form

$$C_D(x) = \gamma_D + \delta_D \Phi(x) \tag{14}$$

with δ_D independent of x and $\Phi(x)$ independent of D. We normalize by taking $\delta_5 = 1$. Then the first few values of $\Phi(x) = C_5^0(x)$ can be read off from the first column of the above table, while the first few values of δ_D are as follows:

D	5	8	12	13	17	20	21	24	28	29
δ_D	1	−2	12	−11	−2	8	42	−108	112	−33

Equation (14) says that the functions $\{C_D(x) \mid D > 0\}$ span a two-dimensional vector space, while Theorems 1 and 2 said that the vector spaces spanned by $\{A_D(x)\}$ and $\{B_D(x)\}$ were one-dimensional. The answer in the general situation, when the exponent 1, 3 or 5 is replaced by an arbitrary odd number (which for later purposes we denote $k-1$) is given by:

Theorem 3. *For every positive even integer k and every positive non-square discriminant D, the function $F_{k,D} : \mathbb{R} \to \mathbb{R}$ defined by*

$$F_{k,D}(x) = \sum_{\substack{a,b,c \in \mathbb{Z},\, a<0,\\ b^2-4ac=D}} \max(0, (ax^2+bx+c)^{k-1}) \tag{15}$$

is a linear combination, with coefficients depending on D and k, of a finite collection of functions depending only on k.

In other words, the functions $F_{k,D}$ for k fixed and D varying span a space of finite dimension. In fact the dimension is $[k/6]+1$, as we will see later.

5. What is going on?

The results of §4 naturally suggest several questions, in particular the following:

• Why are the functions $F_{2,D}(x) = A_D(x)$ and $F_{4,D}(x) = B_D(x)$ constant but the function $F_{6,D}(x) = C_D(x)$ is not?

• Is there a relationship between the numbers γ_D and δ_D and the value of $L_D(-5)$ analogous to the relationships between α_D and $L_D(-1)$ and between β_D and $L_D(-3)$ described earlier?

• What is the nature of the function $\Phi(x)$ in (14)? And what about its coefficient δ_D?

We describe the answers to these questions briefly here. (The proofs will be given later.) We begin with the question about the relationship with special values of L-series. Note that the function $F_{k,D}(x)$ defined in (15) is periodic with period 1, so it has a well-defined average value.

Theorem 4. *For all k and D the average value of $F_{k,D}(x)$ equals $\frac{\zeta_D(1-k)}{2\zeta(1-2k)}$.*

We will prove this in §8. Accepting it for the moment, we can deduce from (14) that the numbers γ_D and δ_D are related to $\zeta_D(-5)$ by

$$\gamma_D + \lambda\delta_D = \frac{\zeta_D(-5)}{2\zeta(-11)} = \frac{16380}{691}\zeta_D(-5) = -\frac{65}{691}L_D(-5),$$

where λ is the average value of the periodic function $\Phi(x)$. Applying this to $D = 5$, 8 and 12, for instance, and using the known values of $L_D(-5)$, we find the three equations

$$2 + \lambda = \frac{1742}{691}, \quad 35 - 2\lambda = \frac{23465}{691}, \quad 310 + 12\lambda = \frac{218530}{691}$$

which agree (fortunately!) and all give the value $\lambda = \frac{360}{691}$. We can therefore rewrite (14) in the alternative form

$$C_D(x) = -\frac{65}{691}L_D(-5) + \delta_D\Phi_0(x), \tag{16}$$

where $\Phi_0(x) = \Phi(x) - \lambda$ is a periodic even function with average value 0.

Our next object is to identify $\Phi_0(x)$. A graph of this function is shown in Figure 2. The graph looks very smooth, but this is misleading. In fact, we have:

Fifth Surprise. *The function $\Phi(x)$ is not C^∞, or even C^{10}.*

Proof. From the definition of $C_D(x)$ it follows that $C_D(x)$, and hence also $\Phi(x)$, has the property that its value at a rational argument $x = p/q$ has the form m/q^{10} with $m \in \mathbb{Z}$. We claim that the only smooth functions $f(x)$ which can have this property are polynomials of degree ≤ 10. (This will imply our statement, since $\Phi(x)$ is non-constant and periodic and hence certainly not a polynomial.) To see this, we note that the 10th derivative of $f(x)$ is given by the formula

$$f^{(10)}(x) = \lim_{Q\to\infty}\left(\sum_{n=0}^{10}(-1)^n Q^{10}\binom{10}{n} f\left(x + \frac{n}{Q}\right)\right).$$

If $x = p/q$ and we let Q tend to infinity through multiples of q, then every term in the sum is an integer, so $f^{10}(x)$ is integral at all rational arguments and hence, if continuous, is constant. □

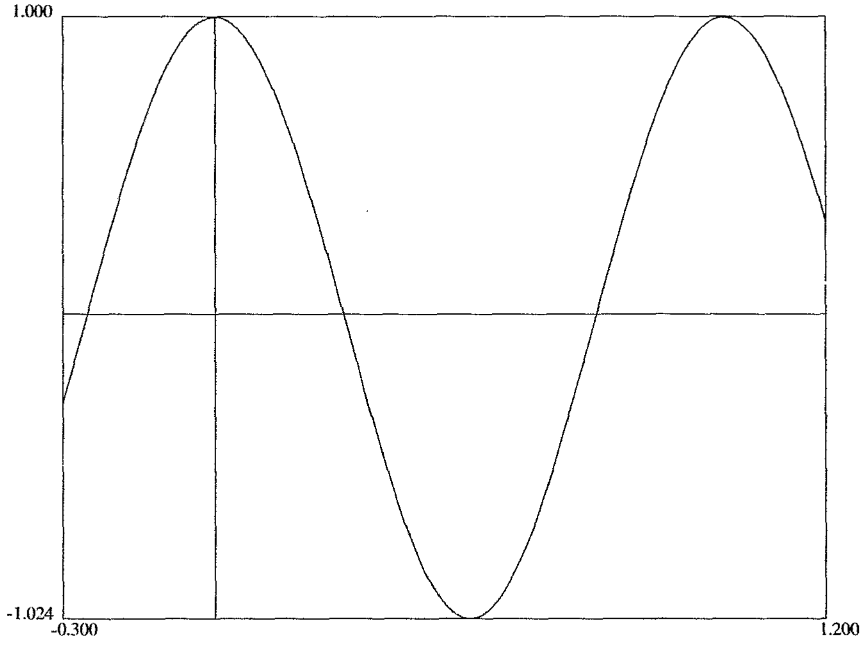

Figure 2

To recognize the mysterious function $\Phi(x)$ or its renormalized variant $\Phi_0(x)$, we could try looking at some special values. But a few specimens were already listed above and they were not very enlightening. A better approach turns out to be to look at the Fourier expansion. Since $\Phi_0(x)$ is even and periodic and has average value 0, it has a Fourier cosine expansion of the form

$$\Phi_0(x) = \sum_{n=1}^{\infty} c_n \cos 2\pi n x$$

with real coefficients c_n. Computing numerically, we find, e.g.,

$$c_1 = 1.01067994278313526\dots.$$

This is again not too revealing, but if we renormalize the further coefficients by dividing by c_1, then something interesting does happen: the quotients c_n/c_1 are (to high numerical precision) rational numbers, the first three values being

$$\frac{c_2}{c_1} = -\frac{3}{256}, \quad \frac{c_3}{c_1} = \frac{28}{19683}, \quad \frac{c_4}{c_1} = -\frac{23}{65536}.$$

The denominator of c_n/c_1 is in each case a divisor of n^{11}, and multiplying by n^{11} to get integers we obtain the values

$$2^{11}\frac{c_2}{c_1} = -24, \quad 3^{11}\frac{c_3}{c_1} = 252, \quad 4^{11}\frac{c_4}{c_1} = -1472,$$

which any number-theorist recognizes at once: they are the first values of the *Ramanujan tau-function*, defined by the expansion

$$q\prod_{n=1}^{\infty}(1-q^n)^{24}=\sum_{n=1}^{\infty}\tau(n)q^n. \tag{17}$$

The meaning of this function is also well-known: the expression occurring on the right-hand side of (17) is the Fourier expansion of a cuspidal modular form of weight 12 on the full modular group $SL(2,\mathbb{Z})$, and 12 is the lowest weight for which such a modular form exists. (See §11 for a very brief review of the theory of modular forms.) So we can finally give the "real" answer to the first and most puzzling of the questions above, namely:

The reason why the function $F_{k,D}(x)$ behaves differently for $k\geq 6$ than it does for $k=2$ and $k=4$ is that there are cusp forms of weight $2k$ on $SL(2,\mathbb{Z})$ for $k\geq 6$ but not for $k<6$.

Thus, starting with our totally elementary "sums of powers of quadratic polynomials" construction, we have been led inexorably to the theory of modular forms, one of the cornerstones of modern number theory. The reason for this connection will be explained in Part III. First, however, we give the proofs of Theorems 1–3, which are completely elementary, and discuss some further refinements and variants of these theorems.

Part II. Proofs, refinements, complements

6. Proofs of Theorems 1–3

Let $A(x)$ be the sum defined by (1). We ignore the question of convergence for the moment, either assuming it or else considering only $x\in\mathbb{Q}$, where the sum is finite. We calculate:

$$x^2A(1/x)-A(x)=\sum_{a<0}\max(0,a+bx+cx^2)-\sum_{a<0}\max(0,ax^2+bx+c)$$

(all summations extend over $(a,b,c)\in\mathbb{Z}^3$ with $b^2-4ac=5$)

$$=\sum_{c<0}\max(0,ax^2+bx+c)-\sum_{a<0}\max(0,ax^2+bx+c)$$

(here we have interchanged a and c in the first sum)

$$=\sum_{c<0<a}\max(0,ax^2+bx+c)-\sum_{a<0<c}\max(0,ax^2+bx+c)$$

(because the terms with a and c both negative in the two previous sums cancel; note that ac can never vanish since 5 is not a square)

$$=\sum_{a>0>c}\max(0,ax^2+bx+c)+\sum_{a>0>c}\min(0,ax^2+bx+c)$$

(here we have replaced a, b and c in the second sum by their negatives)

$$= \sum_{a>0>c} (ax^2 + bx + c) \tag{18}$$

(because $\max(X,0) + \min(X,0) = X$ for any $X \in \mathbb{R}$)

$$= (x^2 + x - 1) + (x^2 - x - 1)$$

(because the only two polynomials $ax^2 + bx + c$ with integer coefficients, discriminant 5, and $a > 0 > c$ are $x^2 \pm x - 1$)

$$= 2x^2 - 2$$

(this step is left to the reader).

It follows that the function $A^0(x) := A(x) - 2$ satisfies the functional equation $x^2 A^0(1/x) = A^0(x)$ as well as the obvious invariance property $A^0(x+1) = A^0(x)$, and since $A^0(0) = 0$ and every rational number can be reduced to 0 by a finite number of iterations of the transformations $x \mapsto x \pm 1$ and $x \mapsto 1/x$ (Euclidean algorithm), this shows that $A^0(x) = 0$ for all $x \in \mathbb{Q}$. This proves our "First surprise" for rational values of x. If we assume that $A(x)$ is continuous (which can in fact be proved in an elementary way), then we deduce that $A(x) = 2$ for all real x. Without this assumption, we at least deduce that the sum in (1) is always convergent and that $A(x) \le 2$ for all $x \in \mathbb{R}$. (*Proof.* If for some value of x the sum either diverged or had a value bigger than 2, then, since all summands are positive, there would be finitely many quadratic functions Q whose sum at x already exceeded 2, and the sum of their values at a sufficiently nearby rational number would also exceed 2, a contradiction.)

Now, to prove Theorem 1, we must replace 5 by an arbitrary non-square discriminant D. The argument up to (18) remains the same and gives

$$x^2 A_D(1/x) - A_D(x) = \sum_{\substack{b^2-4ac=D \\ a>0>c}} (ax^2 + bx + c), \tag{19}$$

but now we no longer know explicitly what forms $[a,b,c]$ contribute to this sum. However, there are only finitely many (independent of x), because the equation $D = b^2 - 4ac = b^2 + 4|ac|$ bounds all three coefficients, so the right-hand side of (19) is a quadratic polynomial in x. Moreover, it is easy to see what this polynomial is: the coefficient of x^2 is the number defined by the right-hand side of equation (3), the constant term is the negative of this, and the coefficient of x is 0 (because for every form contributing to the sum we also have a form with the opposite value of b). In other words, we have

$$x^2 A_D(1/x) - A_D(x) = \alpha_D(x^2 - 1) \tag{20}$$

for every D. Now the rest of the argument goes just like in the special case $D = 5$: the function $A_D^0(x) = A_D(x) - \alpha_D$ vanishes for $x = 0$ and satisfies the functional equations $A_D^0(x+1) = A_D^0(x)$ and $x^2 A_D^0(1/x) = A_D^0(x)$, and from this it follows as before that $A_D^0(x)$ vanishes for all rational arguments and therefore, assuming continuity, for all real arguments.

We now turn to $B(x)$. The same calculation as for $A(x)$, but with all summands cubed, gives

$$x^6B(1/x)-B(x)=\sum_{\substack{b^2-4ac=5\\ a>0>c}}(ax^2+bx+c)^3=(x^2+x-1)^3+(x^2-x-1)^3=2x^6-2,$$

and from this we deduce exactly as before that the function $B^0(x)=B(x)-2$ vanishes at all rational and hence at all real values of x. (This time there is no problem of continuity, since the series defining $B(x)$ converges uniformly.) For general D, however, things are not quite so obvious. The calculation up to (18) is unchanged and gives

$$x^6B_D(1/x)-B_D(x)=P_{4,D}(x):=\sum_{\substack{b^2-4ac=D\\ a>0>c}}(ax^2+bx+c)^3. \tag{21}$$

It is clear that the right-hand side is a polynomial of degree 6, even (replace b by $-b$ in the sum), and anti-invariant under $P(x)\mapsto x^6P(1/x)$, so it has the form

$$P_{4,D}(x)=\beta_Dx^6+\beta'_Dx^4-\beta'_Dx^2-\beta_D \tag{22}$$

with β_D given by equation (10) and some $\beta'_D\in\mathbb{Z}$. But to complete the argument, we have to know that β'_D vanishes. For any particular value of D this can be checked by numerical calculation, e.g. for $D=8$, 12 and 13 we find

$$\begin{aligned}
P_{4,8}(x)&=(x^2-2)^3+(2x^2-1)^3+(x^2+2x-1)^3+(x^2-2x-1)^3\\
&=11x^6-11,\\
P_{4,12}(x)&=(x^2-3)^3+(3x^2-1)^3+(2x^2+2x-1)^3\\
&\quad+(2x^2-2x-1)^3+(x^2+2x-2)^3+(x^2-2x-2)^3\\
&=46x^6-46,\\
P_{4,13}(x)&=(x^2+x-3)^3+(x^2-x-3)^3+(3x^2+x-1)^3\\
&\quad+(3x^2-x-1)^3+(x^2+3x-1)^3+(x^2-3x-1)^3\\
&=58x^6-58,
\end{aligned}$$

but it is not immediately clear why this should always happen. The argument that it is so is the key to the whole phenomenon discussed in the paper and, as we will see in Part III, also to the relationship with the theory of modular forms. From equation (21) and the fact that $B_D(x)$ is even and periodic, we deduce

$$\begin{aligned}
&P_{4,D}(x+1)-P_{4,D}(x)\\
&\quad=(x+1)^6B_D\Big(\frac{1}{x+1}\Big)-B_D(x+1)-x^6B_D\Big(\frac{1}{x}\Big)+B_D(x)\\
&\quad=(x+1)^6B_D\Big(1-\frac{1}{x+1}\Big)-x^6B_D\Big(\frac{1}{x}+1\Big)=x^6P_{4,D}\Big(\frac{1}{x}+1\Big).
\end{aligned}\tag{23}$$

A direct calculation shows that the polynomial x^6-1 satisfies the functional equation $P(x+1)-P(x)=x^6P(1+1/x)$ but that the polynomial x^4-x^2 does not, so from (22) and (23) we deduce that $\beta'_D=0$, and we conclude as before.

Turning now to the 5th powers, we see clearly what changes: the calculation for $C(x)$ is just the same as for $A(x)$ and $B(x)$, but this time we find

$$(x^2+x-1)^5+(x^2-x-1)^5 = 2x^{10}+10x^8-30x^6+30x^4-10x^2-2 \neq \gamma_5(x^{10}-1),$$

so $C(x)-\gamma_5$ cannot vanish identically. For general D we find

$$x^{10}C_D(1/x) - C_D(x) = P_{6,D}(x) := \sum_{\substack{b^2-4ac=D \\ a>0>c}} (ax^2+bx+c)^5$$

where $P_{6,D}(x)$ has the form

$$P_{6,D}(x) = \gamma_D(x^{10}-1) + \gamma'_D(x^8-x^2) + \gamma''_D(x^6-x^4) \tag{24}$$

with γ_D as in (13) and some coefficients γ'_D and γ''_D. The same calculation as in (23) shows that $P_{6,D}(x+1) = P_{6,D}(x) + x^{10}P_{6,D}(1+1/x)$, and substituting (24) into this we find that $\gamma''_D = -3\gamma'_D$ or (setting $\delta_D := \gamma'_D/10$, which is always an integer)

$$P_{6,D}(x) = \gamma_D(x^{10}-1) + 10\delta_D x^2(x^2-1)^3 = \gamma_D(x^{10}-1) + \delta_D(P_{6,5}(x) - 2(x^{10}-1)),$$

and by the now familiar argument this implies equation (14) with $\Phi(x) := C_5^0(x)$.

The argument obviously works for all k. For k and D as in Theorem 3, we set

$$P_{k,D}(x) = \sum_{\substack{b^2-4ac=D \\ a>0>c}} (ax^2+bx+c)^{k-1} \quad \in \mathbb{Z}[x]. \tag{25}$$

Then the same calculation as before shows that $F_{k,D}$ and $P_{k,D}$ are related by

$$x^{2k-2}F_{k,D}(1/x) - F_{k,D}(x) = P_{k,D}(x) \tag{26}$$

and that $P_{k,D}$ belongs to the (finite-dimensional) vector space $\mathfrak{W}^+_{2k-2}$ of all polynomials satisfying the functional equation $P(x+1) = P(x) + x^{2k-2}P(1+1/x)$. It follows that $F_{k,D}(x)$ is a linear combination of finitely many (more precisely: $\dim \mathfrak{W}^+_{2k-2}$) functions of x depending only on k, the coefficients being simply the coordinates of the polynomial $P_{k,D}$ with respect to some fixed basis of the space $\mathfrak{W}^+_{2k-2}$. This proves Theorem 3, with a relatively explicit description of the linear combination whose existence it asserts.

7. Modifications when the discriminant is a square

When D is a square, Theorems 1–3 with the definitions given earlier do not hold. It turns out that they remain true if we modify the definitions. We assume throughout the section that $D = m^2$ is a positive square, and denote by $A^*_D(x)$, $B^*_D(x)$ and $F^*_{k,D}(x)$ the expressions on the right-hand sides of equations (2), (9) and (15). We will see that the "right" definitions of A_D, B_D and $F_{k,D}$ in this case are the sums of these expressions with simple correction terms. Let us start as in Part I with some experimental data. We consider the simplest case $D = 1$ and compute the sums in (2) and (9) for some simple rational values of x. The functions defined by (2) and (9) are periodic and even, so we can restrict to values between 0 and 1/2.

If x has the form $1/n$ with n positive, then one can show without too much trouble that the quadratic forms $Q=[a,b,c]$ of discriminant 1 with $a<0$ and $Q(x)>0$ are given by $Q=[-r,1,0]$ with r ranging from 1 to $n-1$. The corresponding values of $Q(x)$ are $(n-r)/n^2$, so we find

$$A_1^*\Big(\frac{1}{n}\Big)=\sum_{r=1}^{n}\Big(\frac{1}{n}-\frac{r}{n^2}\Big)=\frac{n-1}{2n},\quad B_1^*\Big(\frac{1}{n}\Big)=\sum_{r=1}^{n}\Big(\frac{1}{n}-\frac{r}{n^2}\Big)^3=\frac{(n-1)^2}{4n^4}.$$

For the other fractions in $[0,\frac{1}{2}]$ with denominator less than 8, we find the values

Q	$Q(2/5)$
$[-1,1,0]$	$6/25$
$[-2,1,0]$	$2/25$
$[-6,5,-1]$	$1/25$

Q	$Q(2/7)$
$[-1,1,0]$	$10/49$
$[-2,1,0]$	$6/49$
$[-3,1,0]$	$2/49$
$[-12,7,-1]$	$1/49$

Q	$Q(3/7)$
$[-1,1,0]$	$12/49$
$[-2,1,0]$	$3/49$
$[-6,5,-1]$	$2/49$
$[-10,9,-2]$	$1/49$

and hence

$$A_1^*(\tfrac{2}{5})=\tfrac{9}{25},\quad B_1^*(\tfrac{2}{5})=\tfrac{9}{625},\quad A_1^*(\tfrac{2}{7})=\tfrac{19}{49},$$
$$B_1^*(\tfrac{2}{7})=\tfrac{25}{2401},\quad A_1^*(\tfrac{3}{7})=\tfrac{18}{49},\quad B_1^*(\tfrac{3}{7})=\tfrac{36}{2401}.$$

Looking at all this data we are led to conjecture the following formula for $A_1^*(x)$ and $B_1^*(x)$:

Proposition. *Let x be a real number between 0 and 1. Then*

$$\sum_{\substack{\mathrm{disc}(Q)=1\\ a<0<Q(x)}} Q(x)=\tfrac{1}{2}(x^2-x+1)-\tfrac{1}{2}\kappa(x),\qquad \sum_{\substack{\mathrm{disc}(Q)=1\\ a<0<Q(x)}} Q(x)^3=\tfrac{1}{4}x^2(x-1)^2,\quad (27)$$

where $\kappa(x)=1/q^2$ if $x=p/q$ with $(p,q)=1$ and $\kappa(x)=0$ if x is irrational.

We will give a direct (and quite amusing) proof of this in §10 using continued fractions. For now we continue to more general square discriminants. We want to find a modification of $A_D(x)$, $B_D(x)$ and more generally $F_{k,D}(x)$ such that Theorems 1–3 remain valid. Moreover, in order for the formulas (7) and (11) to remain true for $D=1$ (the only square discriminant which is fundamental; we have not yet defined $L_D(s)$ for D not fundamental), the values of the constant functions $A_1(x)$ and $B_1(x)$ should equal $\alpha_1=-5L_1(-1)=-5\zeta(-1)=\frac{5}{12}$ and $\beta_1=L_1(-3)=\zeta(-3)=\frac{1}{120}$. In view of (27) this means that we have to define

$$A_1(x)=A_1^*(x)-\frac{1}{2}\mathbb{B}_2(x)+\frac{1}{2}\kappa(x),\quad B_1(x)=B_1^*(x)-\frac{1}{4}\mathbb{B}_4(x)$$

for $0\le x\le 1$, where

$$\mathbb{B}_2(x)=x^2-x+\frac{1}{6},\quad \mathbb{B}_4(x)=x^4-2x^3+x^2-\frac{1}{30}$$

denote the 2nd and 4th Bernoulli polynomials. (We use the letter $\mathbb{B}$ because $B_D(x)$ has already been assigned a different meaning.) For x not between 0 and 1, we must replace $\mathbb{B}_k(x)$ ($k = 2, 4$) by $\bar{\mathbb{B}}_k(x) = \mathbb{B}_k(x - [x])$, the periodic version of the Bernoulli polynomial, in order to achieve the desired periodicity. For larger square values of D we find experimentally that the right formulas are

$$\begin{aligned} A_{m^2}(x) &= A^*_{m^2}(x) - \frac{1}{2}\bar{\mathbb{B}}_2(mx) + \frac{1}{2}m^2\kappa(x), \\ B_{m^2}(x) &= B^*_{m^2}(x) - \frac{1}{4}\bar{\mathbb{B}}_4(mx), \end{aligned} \tag{28}$$

and similarly for higher powers

$$F_{k,m^2}(x) = F^*_{k,m^2}(x) - \frac{1}{k}\bar{\mathbb{B}}_k(mx) \quad (k > 2). \tag{29}$$

Claim. With these definitions, Theorems 1–3 remain valid also for square values of D.

Proof. We indicate the argument briefly. The main point is to prove that the relation (26) remains true for $D = m^2$ if we define $F_{k,D}$ by (29) and $P_{k,D}$ by

$$P_{k,m^2}(x) = P^*_{k,m^2}(x) + \frac{1}{k}\left(\mathbb{B}_k(mx) - x^{2k-2}\mathbb{B}_k\left(\frac{m}{x}\right)\right), \tag{30}$$

where $P^*_{k,D}(x)$ denotes the polynomial on the right-hand side of (25). Indeed, the same argument as in the derivation of (18) shows that

$$\begin{aligned} x^{2k-2}F^*_{k,m^2}(1/x) - F^*_{k,m^2}(x) &= \sum_{c<0\le a} \max(0, Q(x)^{k-1}) - \sum_{a<0\le c} \max(0, Q(x)^{k-1}) \\ &= P^*_{k,m^2}(x) + \sum_{c<0=a} \max(0, Q(x)^{k-1}) - \sum_{a<0=c} \max(0, Q(x)^{k-1}) \\ &= P^*_{k,m^2}(x) + \sum_{n\ge 1}\{\max(0, |mx| - n)^{k-1} - \max(0, |mx| - nx^2)^{k-1}\}. \end{aligned}$$

Now applying to $t = |mx|$ and $t = |m/x|$ the identity

$$\sum_{n\ge 1} \max(0, t-n)^{k-1} = \frac{\mathbb{B}_k(t) - \bar{\mathbb{B}}_k(t)}{k} \quad (t \ge 0)$$

[*Proof*: both sides vanish if $0 \le t < 1$ and increase by t^{k-1} when we replace t by $t+1$], we find

$$\begin{aligned} &x^{2k-2}F^*_{k,m^2}(1/x) - F^*_{k,m^2}(x) \\ &\qquad = P^*_{k,m^2}(x) + \frac{\mathbb{B}_k(mx) - \bar{\mathbb{B}}_k(mx)}{k} - x^{2k-2}\frac{\mathbb{B}_k(m/x) - \bar{\mathbb{B}}_k(m/x)}{k}. \end{aligned}$$

In view of the definitions (29) and (30) this is the desired equation (26), at least for $k > 2$, but in fact also for $k = 2$ because the extra term $\kappa(x)$ in (28) satisfies $x^2\kappa(1/x) = \kappa(x)$.

To complete the proof, we have to consider the extent to which equation (26) characterizes $F_{k,D}$. In proving the constancy of (for instance) $A_D(x)$ for non-square D, we used that the coefficient of x^2 in (19) was equal to the same number α_D, given by (3), as the value of $A_D(0)$, so that the function $A_D^0 = A_D - \alpha_D$ was not only invariant (up to automorphy factors) under $SL(2,\mathbb{Z})$, but also vanished at 0 and hence at all rational numbers. Here this would not be true without the second correction term $\frac{1}{2}D\kappa(x)$ in (28). Indeed, denoting by α_D^* the expression defined by the right-hand side of (3) when D is a square, we find from (30) that $P_{2,D}(x)$ equals $\alpha_D(x^2-1)$ with $\alpha_D = \alpha_D^* - \frac{1}{12} + \frac{1}{2}D$, so $A_D(0)$, if $A_D(x)$ is to be constant, must have this value α_D, and this is achieved precisely by including the second correction term in (28). In general, the freedom on rational numbers of a function F on $\mathbb{R}$ satisfying $F(x+1) = F(x) = x^{2n}F(1/x)$ is given precisely by multiples of the function $\kappa(x)^n$, since this function satisfies the given functional equations and the values of such an F at all rational points are determined by its value at a single point. But for $k > 2$ the function $F_{k,D}(x)$ is continuous (because the series defining $F_{k,D}^*$ converges uniformly and $\mathbb{B}_k(x)$ is also continuous), so the multiple of the (a priori possible, but discontinuous) extra term $\kappa(x)^{k-1}$ must be zero. □

We have shown in particular that the equations $A_D(x) = \alpha_D$, $B_D(x) = \beta_D$ and (14) (with the same function $\Phi(x)$ as before) remain true for square values of D. Here is a table of the numbers α_D, β_D, γ_D and δ_D for some small squares (including the case $D = 0$ for which the function $F_{k,D}$ will be defined in the next section):

D	0	1	4	9	16	25	36
$12\alpha_D$	$-\frac{1}{2}$	5	35	125	275	605	875
$120\beta_D$	$\frac{1}{2}$	1	121	2161	15481	78001	261481
$252\gamma_D$	$-\frac{1}{2}$	-1	251	16631	389339	4590935	33565139
$120\delta_D$	0	1	-56	9	-704	1705	-504

8. The average value of $F_{k,D}(x)$

We now turn to the evaluation of the average value of the even periodic function $F_{k,D}$. The answer was stated in §5 as Theorem 4 in terms of a special value of the L-function $L_D(s)$, but so far we have defined this only for fundamental discriminants D. We first explain how the definition is extended to arbitrary positive discriminants, and then describe the (very simple) proof of the theorem.

We discussed in §2 the definitions and main properties of the functions $L_D(s)$ and $\zeta_D(s) = \zeta(s)L_D(s)$ defined by equation (4) when D is the discriminant of a real quadratic field K. In particular, the values of these functions at negative odd arguments $s = 1 - k$ ($k \geq 2$ even) are rational numbers, and the values at the corresponding positive arguments $s = +k$ are rational multiples of a power of π times the square root of D, because of the functional equations (5). In [Co3], Henri Cohen defined a number-theoretical function $H(k, D)$ for all positive even k and all nonnegative integers D satisfying the two properties:

(i) $H(k, D) = L_D(1-k)$ when D is 1 or the discriminant of a real quadratic field, and
(ii) $\sum_{D=0}^{\infty} H(k, D)q^D$ is a modular form of weight $k + \frac{1}{2}$ on $\Gamma_0(4)$.

(The definition of a modular form of half-integral weight will be recalled in §14.) The number $H(k, D)$ is defined as $\zeta(1-2k)$ if $D = 0$ and otherwise as a certain multiple of $L_{D_0}(1-k)$, where D_0 is the discriminant of the field $\mathbb{Q}(\sqrt{D})$. More specifically, if D is congruent to 2 or 3 modulo 4, then $H(k, D) = 0$, while if D is congruent to 0 or 1 modulo 4, then $D = D_0 f^2$ for some positive integer f and $H(k, D)$ is defined as the product of $H(k, D_0) = L_{D_0}(1-k)$ with a simple Euler factor depending on the prime factors of f, the simplest case being

$$H(k, D_0p^2) = \Bigl(p^{2k-1} - \Bigl(\frac{D_0}{p}\Bigr)p^{k-1} + 1\Bigr)H(k, D_0) \qquad (D_0 \text{ fundamental, } p \text{ prime}). \tag{31}$$

In [Z3], I gave intrinsic definitions (i.e., not depending on the representation $D = D_0 f^2$) of two Dirichlet series $L_D(s)$ and $\zeta_D(s)$ for every D, satisfying the functional equations (5) and (6) and such that the special value of $L_D(s)$ at $s = 1-k$ agrees with Cohen's function $H(k, D)$ for every positive even integer k and positive discriminant D. The definition of $\zeta_D(s)$ is

$$\zeta_D(s) = \sum_{Q \in \mathfrak{Q}_D/\Gamma} \; \sum_{\substack{(x,y) \in \mathbb{Z}^2/\Gamma_Q \\ Q(x,y)>0}} \frac{1}{Q(x,y)^s}, \tag{32}$$

where $\mathfrak{Q}_D$ denotes the set of all $Q = [a, b, c] \in \mathbb{Z}^3$ with discriminant $b^2 - 4ac = D$ (now thought of homogeneously as quadratic forms $Q(x, y) = ax^2 + bxy + cy^2$ instead of quadratic functions $Q(x) = ax^2 + bx + c$), Γ is the group $PSL(2, \mathbb{Z})$, acting on quadratic forms in the usual way, and $\Gamma_Q = \{\gamma \in \Gamma \mid Q \circ \gamma = Q\}$ denotes the stabilizer of Q in Γ, which for D positive and non-square is always an infinite cyclic group, its generator corresponding to the basic solution of Pell's equation or to the fundamental unit of the quadratic field $\mathbb{Q}(\sqrt{D})$. If D is the discriminant of a real quadratic field K, then the first sum corresponds to the sum over the ideal classes of K and the second sum equals $\sum N(\mathfrak{a})^{-s}$ with $\mathfrak{a}$ running over the integral ideals in a single ideal class, so the new definition coincides with (4). If D is a positive non-square, then $\zeta_D(s)$ is a simple multiple of $\zeta_{\mathbb{Q}(\sqrt{D})}(s)$ and the function $L_D(s) := \zeta_D(s)/\zeta(s)$ is holomorphic and agrees at negative arguments $s = 1-k$ with Cohen's $H(k, D)$. If $D = 1$, then $\zeta_D(s) = \zeta(s)^2$ and $L_D(s) = \zeta(s)$, and finally for $D = 0$ one has $\zeta_D(s) = \zeta(s)\zeta(2s-1)$ and $L_D(s) = \zeta(2s-1)$.

Using this definition, we can give quite an easy proof of Theorem 4 for all discriminants $D > 0$. We first observe that the double summation over $Q \in \mathfrak{Q}_D/\Gamma$

and $(x,y)\in\mathbb{Z}^2/\Gamma_Q$ in (32) can be combined into a single sum:

$$\zeta_D(s)=\sum_{\substack{(Q,(x,y))\in(\mathfrak{Q}_D\times\mathbb{Z}^2)/\Gamma\\ Q(x,y)>0}}\frac{1}{Q(x,y)^s}.$$

Now observing that every non-zero element of $\mathbb{Z}^2$ is $SL(2,\mathbb{Z})$-equivalent to a unique point of the form $(m,0)$ with $m\geq 1$ and that the stabilizer of $(m,0)$ is the group Γ_∞ generated by $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$, we can rewrite this as

$$\zeta_D(s)=\sum_{m=1}^{\infty}\sum_{\substack{Q\in\mathfrak{Q}_D/\Gamma_\infty\\ Q(m,0)>0}}\frac{1}{Q(m,0)^s}=\zeta(2s)\sum_{\substack{Q\in\mathfrak{Q}_D/\Gamma_\infty\\ Q(1,0)>0}}\frac{1}{Q(1,0)^s}=\zeta(2s)\sum_{n=1}^{\infty}\frac{N_D(n)}{n^s},$$

where $N_D(n)$ denotes the number of integers b modulo $2n$ with $b^2\equiv D\ (\mathrm{mod}\,4n)$. (The last step is obtained by writing $Q=[n,b,(D-b^2)/4n]$ with $b^2\equiv D\ (\mathrm{mod}\,4n)$ and noting that the action of Γ_∞ corresponds simply to shifting b by multiples of $2n$.) On the other hand, the average value of $F_{k,D}(x)$ is given by

$$\langle F_{k,D}\rangle_{\mathrm{av}}=\int_0^1 F_{k,D}(x)dx=\sum_{\substack{Q=[a,b,c]\in\mathfrak{Q}_D/\Gamma_\infty\\ a<0}}\beta_k(Q),$$

where $\beta_k(Q)=\int_{-\infty}^{\infty}\max(0,Q(x))^{k-1}dx$. Making the substitution $x=\frac{-b+t\sqrt{D}}{2a}$, we find

$$\beta_k(Q)=c_kD^{k-1/2}|a|^{-k},\quad c_k=\frac{1}{2^{2k-1}}\int_{-1}^{1}(1-t^2)^{k-1}dt=\frac{1}{2^{2k-1}}\frac{\Gamma(k)\Gamma(\frac12)}{\Gamma(k+\frac12)},$$

and hence, since the number of $Q\in\mathfrak{Q}_D/\Gamma_\infty$ with first coefficient $a=-n<0$ is precisely $N_D(n)$,

$$\langle F_{k,D}\rangle_{\mathrm{av}}=c_kD^{k-1/2}\sum_{n=1}^{\infty}\frac{N_D(n)}{n^k}=c_kD^{k-1/2}\frac{\zeta_D(k)}{\zeta(2k)}=\frac12\frac{\zeta_D(1-k)}{\zeta(1-2k)},$$

where the last equality follows from the functional equation (5). This completes the proof of Theorem 4.

If D is a square and $F_{k,D}$ is defined by (29), then Theorem 4 still remains true, since the above calculation shows that the average value of $F^*_{k,D}(x)$ has the expected value and the correction term $-\bar{\mathbb{B}}_k(mx)/k$ in (29) (as well, of course, as the second correction term $\kappa(x)$ in (28), which is non-zero only on a set of measure 0) has the average value zero. Finally, for $D=0$ we simply *define* $F_{k,D}(x)$ as the constant function

$$F_{k,0}(x)=\frac12\zeta(1-k)\quad(\text{all }x\in\mathbb{R}).\tag{33}$$

(This definition will be needed in §14, where it will be shown that the numbers $F_{k,D}(x)$ are the Fourier coefficients of a modular form of weight $k+\frac12$ for every $x\in\mathbb{R}$.) Then the average value of $F_{k,0}$ is also $\frac12\zeta(1-k)$, and this agrees with Theorem 4 because $\zeta_0(s)=\zeta(s)\zeta(2s-1)$.

9. Sums over quadratic forms in an equivalence class

For each discriminant D, the set $\mathfrak{Q}_D$ of quadratic functions $[a,b,c]$ of discriminant D decomposes into a finite number of orbits $\mathcal{A}_1,\ldots,\mathcal{A}_h$ of classes under the action of the group $\Gamma = PSL(2,\mathbb{Z})$. When D is the discriminant of a real quadratic field K, for instance, h is equal to the class number of K, the correspondence between ideal classes and Γ-equivalence classes of forms being given by assigning to each ideal class the class of the norm form of an ideal belonging to it. It is reasonable to ask whether the construction studied in this paper still works when we sum only over the quadratic functions or forms in a single class. The answer is that they do, but only after certain symmetrizations with respect to two involutions on the set $\mathfrak{Q}_D/\Gamma$ of classes. We explain this very briefly. We will consider only the case of non-square discriminants. This is not a serious restriction, since the division into classes when the discriminant equals m^2 is simply a question of congruences modulo m and not very interesting, and has the advantage that we do *not* have to worry about the extra complications which were discussed in §7.

For each class $\mathcal{A} \in \mathfrak{Q}_D/\Gamma$ and each (not necessarily even) integer $k \geq 2$, define

$$F_{k,\mathcal{A}}^*(x) = \sum_{\substack{[a,b,c]\in\mathcal{A} \\ a<0}} \max\bigl(0,(ax^2+bx+c)^{k-1}\bigr). \tag{34}$$

Then $F_{k,D}(x)$ is the sum of the $F_{k,\mathcal{A}}^*$ with $\mathcal{A}$ ranging over $\mathfrak{Q}_D/\Gamma$, but Theorems 1–3 are *not* true for each $F_{k,\mathcal{A}}^*$ separately, e.g., the functions $F_{2,\mathcal{A}}^*(x)$ and $F_{4,\mathcal{A}}^*(x)$ are not constant in general. The reason is that in the derivation of (18) we had to replace the form Q by $-Q$ at one point, and $-Q$ in general belongs to a different Γ-equivalence class than Q. For the same reason, the polynomial $P_{k,\mathcal{A}}^*(x)$ defined by restricting the summation in (25) to $[a,b,c] \in \mathcal{A}$ does not satisfy the functional equation on which the argument of §6 depended. However, if we set $-\mathcal{A} := \{-Q \mid Q \in \mathcal{A}\}$ and

$$\begin{aligned} F_{k,\mathcal{A}}(x) &= F_{k,\mathcal{A}}^*(x) + (-1)^k F_{k,-\mathcal{A}}^*(x), \\ P_{k,\mathcal{A}}(x) &= P_{k,\mathcal{A}}^*(x) + (-1)^k P_{k,-\mathcal{A}}^*(x), \end{aligned} \tag{35}$$

then the same argument as in §6 establishes the identity

$$x^{2k-2}F_{k,\mathcal{A}}(-1/x) - F_{k,\mathcal{A}}(x) = P_{k,\mathcal{A}}(x) \tag{36}$$

in place of (26). (Note the minus sign in the first argument, which was omitted in (26) because the functions $F_{k,D}$ and $P_{k,D}$ are even. We could if we wished restrict here also to even or odd functions by symmetrizing or antisymmetrizing with respect to the second involution $\mathcal{A} \mapsto \mathcal{A}' := \{[a,-b,c] \mid [a,b,c] \in \mathcal{A}\}$ on $\mathfrak{Q}_D/\Gamma$, but there is no particular reason to do so.) Now everything works as in the previous case, except that we have more freedom, since we can now treat both even and odd values of k and both even and odd polynomials. Equation (36) and the periodicity of $F_{k,\mathcal{A}}(x)$ imply that the polynomial $P = P_{k,\mathcal{A}}$ satisfies the *period equation*

$$P(x) = P(1+x) + (x+1)^{2k-2}P\Bigl(\frac{x}{x+1}\Bigr). \tag{37}$$

For $k = 2$, 3, 4, 5 and 7 the only solutions of this equation are the multiples of the polynomial $x^{2k-2} - 1$, so with the same argument as in the proofs of Theorems 1 and 2 we find that each of the functions $F_{k,\mathcal{A}}$ is constant for these values of k. Finally, the value of this constant (and more generally, the average value of the function $F_{k,\mathcal{A}}(x)$ even when it isn't constant) can be calculated by the same method as in §8 and is equal to a simple multiple of the value of the partial zeta function of $\mathcal{A}$ (= the inner sum in (32)) at the argument $s = 1-k$. Observe that the constancy and non-vanishing of (for instance) $F_{4,\mathcal{A}}$ has a nice corollary: for every real number x and every Γ-equivalence class $\mathcal{A}$ there is at least one $Q \in \mathcal{A}$ (infinitely many if x is irrational) such that x lies between the zeros of the quadratic function Q.

The period equation (37) is intimately related to modular forms, as will be recalled in §11 (for instance, the integers $k = 2$, 3, 4, 5, 7 where its solution space is one-dimensional are just the values of $k > 1$ for which there are no cusp forms of weight $2k$ on $SL(2,\mathbb{Z})$), and the functions $F_{k,\mathcal{A}}$ and $P_{k,\mathcal{A}}$ can also be related to modular forms (cf. §13). The polynomials $P_{k,\mathcal{A}}(x)$ also have an interesting application to the problem of classifying "rational period functions" (rational functions satisfying the functional equation (37) for some negative k), as explained in [ChZ].

10. Connections with Diophantine approximation and continued fractions

So far our analysis of the function $A(x)$ and its generalizations was based on the functional equation (26). In this section we briefly discuss a more direct approach, in which we look directly at the quadratic forms of a given discriminant for which a given real number x lies between their two real roots. We will consider only two cases: the case when $D = 1$ and $k = 2$ or 4, where we will give an elementary direct proof of the proposition formulated in §7, and the analysis of the special forms which appeared in the table given in §1 for $D = 5$, $x = 1/\pi$.

We start with the proof of equation (27). Let x be a real number between 0 and 1 and write x as a continued fraction

$$x = \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cfrac{1}{\ddots}}} = [0, n_1, n_2, \ldots] \tag{38}$$

with $n_1, n_2, \ldots \geq 1$. (We suppose for the moment that x is irrational, so that the continued fraction does not terminate.) Define real numbers $\delta_0, \delta_1, \delta_2, \ldots$ inductively by

$$\delta_0 = 1, \quad \delta_1 = x, \quad \delta_{j+1} = \delta_{j-1} - n_j\delta_j \quad (j \geq 1). \tag{39}$$

Note that $n_j = [\delta_{j-1}/\delta_j]$, so that $1 = \delta_0 > \delta_1 > \delta_2 > \ldots > 0$. The successive quotients δ_{j-1}/δ_j are just the continued fractions $[n_j, n_{j+1}, \ldots]$, while the numbers

δ_j themselves are given by $\delta_{j+1} = |p_j - q_j x|$ where $p_j/q_j = [0, n_1, \ldots, n_j]$ is the jth continued fraction approximation to x.

It is an exercise which we leave to the reader to check that the values at x of the quadratic functions $Q = [a, b, c]$ of discriminant 1 with a negative and $Q(x)$ positive are precisely the numbers $\delta_j(\delta_{j-1} - n\delta_j)$ with $j \geq 1$ and $1 \leq n \leq n_j$. It follows that $A_1^*(x)$, the expression given by the first sum in (27), is equal to $\sum_{j\geq 1} \varepsilon_j$, where

$$\varepsilon_j := \sum_{n=1}^{n_j} \delta_j(\delta_{j-1} - n\delta_j) = (n_j\delta_j)(\delta_{j-1} - \tfrac{1}{2}(n_j+1)\delta_j).$$

Using (39), we can rewrite this as

$$\varepsilon_j = \tfrac{1}{2}(\delta_{j-1} - \delta_{j+1})(\delta_{j-1} - \delta_j + \delta_{j+1}) = \tfrac{1}{2}(\delta_{j-1}^2 - \delta_{j+1}^2 + \delta_j\delta_{j+1} - \delta_{j-1}\delta_j),$$

and this implies that the sum $\sum_j \varepsilon_j$ telescopes:

$$\sum_{j=1}^{J} \varepsilon_j = \tfrac{1}{2}(\delta_0^2 - \delta_0\delta_1 + \delta_1^2) - \tfrac{1}{2}(\delta_J^2 - \delta_J\delta_{J+1} + \delta_{J+1}^2). \tag{40}$$

Letting $J \to \infty$ in (40) and observing that δ_J tends rapidly (indeed, exponentially rapidly) to 0, we deduce

$$A_1^*(x) = \sum_{j=1}^{\infty} \varepsilon_j = \tfrac{1}{2}(\delta_0^2 - \delta_0\delta_1 + \delta_1^2) = \tfrac{1}{2}(1 - x + x^2),$$

which is the first of equations (27) in the case when x is irrational. If $x = p/q$ is rational, then the argument is the same except that the continued fraction (38) terminates at some n_J. Then $\delta_J = 1/q$ and $\delta_{J+1} = 0$, so the right-hand side of (40) equals $\frac{1}{2}(1 - x + x^2) - \frac{1}{2}q^{-2}$, proving the first of equations (27) in this case also. For the sums of the cubes the argument is similar, but the calculation is messier: we have $B_1^*(x) = \sum \varepsilon_j'$ with

$$\begin{aligned}\varepsilon_j' = \sum_{n=1}^{n_j} \delta_j^3(\delta_{j-1} - n\delta_j)^3 &= -(n_j\delta_j)^4\left(\tfrac{1}{4}\delta_j^2\right) + (n_j\delta_j)^3\left(\delta_j^2\delta_{j-1} - \tfrac{1}{2}\delta_j^3\right)\\ &\quad - (n_j\delta_j)^2\left(\tfrac{3}{2}\delta_j^2\delta_{j-1}^2 - \tfrac{3}{2}\delta_j^3\delta_{j-1} + \tfrac{1}{4}\delta_j^4\right) + (n_j\delta_j)\left(\delta_j^2\delta_{j-1}^3 - \tfrac{3}{2}\delta_j^3\delta_{j-1}^2 + \tfrac{1}{2}\delta_j^4\delta_{j-1}\right)\\ &= \tfrac{1}{4}\delta_j^2\delta_{j-1}^2(\delta_{j-1} - \delta_j)^2 - \tfrac{1}{4}\delta_{j+1}^2\delta_j^2(\delta_j - \delta_{j+1})^2\end{aligned}$$

and the series again telescopes:

$$\sum_{j=1}^{J} \varepsilon_j' = \tfrac{1}{4}\delta_1^2\delta_0^2(\delta_0 - \delta_1)^2 - \tfrac{1}{4}\delta_{J+1}^2\delta_J^2(\delta_J - \delta_{J+1})^2. \tag{41}$$

Letting $J \to \infty$, we obtain $B_1^*(x) = \frac{1}{4}x^2(x-1)^2$ as claimed. Here the rational values of x do not behave differently since if J is the length of the continued fraction then $\delta_J > \delta_{J+1} = 0$ and hence the second term in (41) vanishes. If we attempted to repeat the same argument for fifth powers, we would simply find an expression which did not telescope, so that we would fail to get a closed formula for the sum $C_1^*(x) = \sum Q(x)^5$. (But it would be pretty hard to see from that point of view that the true reason for this failure is the existence of cusp forms of weight 12 !)

The computation just given, although not completely obvious, was relatively easy because it was possible to describe in closed form the quadratic functions of discriminant 1 with $a < 0 < Q(x)$ in terms of the continued fraction expansion of x. For non-square discriminants it is far less clear how to do this. However, in studying the functions Q arising in the table given in §1 for the special case $D = 5$, $x = 1/\pi$, it became apparent that there was in fact a strong pattern. Specifically, each quadratic function Q which appeared on the list was obtained from the preceding one by applying a fairly simple element of $SL(2,\mathbb{Z})$, and using this observation it was possible to write a computer program which generated all of the Q's very quickly. For instance, the 30th function Q turns out to be the polynomial

$$\begin{aligned} Q(x) = &- 535055621994441675779x^2 \\ &+ 3406269882780961090 55x - 542124689349640 85845 \end{aligned}$$

of discriminant 5, with $Q(1/\pi) \approx 9.31 \times 10^{-10}$. I have not yet been able to find a complete description of all the functions Q on the list, but a partial (empirical!) description is that the functions Q occurring are *all* of the expressions $\delta_j^2 + \delta_j\delta_{j+1} - \delta_{j+1}^2$ and *some* of the expressions $\delta_j^2 - \delta_j\delta_{j+1} - \delta_{j+1}^2$, where the linear forms $\delta_j \in \mathbb{Z} + \mathbb{Z}x$ are defined as at the beginning of this section by the continued fraction expansion of $x = 1/\pi$. If this is true, then the exponential decay of the δ_j explains the rapid convergence of the series $A(x)$.

Part III. The modular connection

11. Periods of modular forms

For an even integer $n > 0$, denote by $\mathfrak{V}_n$ the set of polynomials $P(x)$ of degree $\leq n$ and by $\mathfrak{W}_n$ the subset of polynomials satisfying the period equation

$$P(x) = P(1 + x) + (x + 1)^n P\Big(\frac{x}{x+1}\Big). \tag{42}$$

Note that the group $\Gamma = PSL(2,\mathbb{Z})$ acts on $\mathfrak{V}_n$ by

$$P(x) \mapsto (P|_{-n}\gamma)(x) := (cx + d)^n P\Big(\frac{ax + b}{cx + d}\Big)$$

for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$, and that (42) can be rewritten in the form $P|(1 - T - T') = 0$, where T and T' denote the generators $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ of Γ. (Here we have extended the action of the group Γ to an action of the group ring $\mathbb{Z}[\Gamma]$ in the obvious way.)

It is an (amusing) exercise to check that

$$P \in \mathfrak{W}_n \Longleftrightarrow P|(1+S) = P|(1+U+U^2) = 0,$$

where $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $U = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$ are the standard generators of Γ of order 2 and 3. (*Hint*: We have $T = US$ and $T' = U^2 S$, and if the polynomials $P|(1+S)$ and $P|(1+U+U^2)$ coincide then they must both vanish since $\mathfrak{V}_n^\Gamma = \{0\}$ for $n > 0$.) It is also easy to check that if $\mathfrak{W}_n^+$ (resp. $\mathfrak{W}_n^-$) denotes the subspace of $\mathfrak{W}_n$ consisting of even (resp. odd) polynomials, then $\mathfrak{W}_n = \mathfrak{W}_n^+ \oplus \mathfrak{W}_n^-$ (i.e., the even and odd part of a polynomial belonging to $\mathfrak{W}_n$ again belong to $\mathfrak{W}_n$) and that

$$P \in \mathfrak{W}_n^\pm \Longleftrightarrow P(x+1) = P(x) \pm x^n P(1+1/x);$$

in particular, the notation agrees with the notation $\mathfrak{W}_{2k-2}^+$ used at the end of §6. The basic property of $\mathfrak{W}_n$ is

$$P \in \mathfrak{V}_n, \quad P(x) = F(x) - x^n F(-1/x) \text{ for some periodic } F \\ \Longrightarrow \quad P \in \mathfrak{W}_n. \tag{43}$$

(Here "periodic" means periodic of period 1, i.e. $F(x+1) = F(x)$.) This is precisely the property which we used in §6 and §9 to deduce that $P_{k,D} \in \mathfrak{W}_{2k-2}^+$ and $P_{k,\mathcal{A}} \in \mathfrak{W}_{2k-2}$.

The theory of periods says that there is an intimate connection between the spaces $\mathfrak{W}_n$ and modular forms on Γ. Recall that a modular form of weight k on Γ (k a positive even integer) is a holomorphic function in the complex upper half-plane $\mathfrak{H} = \{z \in \mathbb{C} \mid \Im(z) > 0\}$ satisfying the functional equation

$$f(z) = (f|_k\gamma)(z) := (cz+d)^{-k} f\Big(\frac{az+b}{cz+d}\Big) \tag{44}$$

for all matrices $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ and having a Fourier expansion of the form

$$f(z) = \sum_{n=0}^{\infty} a(n) q^n \quad (q := e^{2\pi i z}). \tag{45}$$

In view of the fact that (45) already forces the T-invariance of f and that T and S generate Γ, it is sufficient to require (44) only for the matrix $\gamma = S$, where it takes on the simpler form $f(-1/z) = z^k f(z)$. We denote by $\mathfrak{M}_k$ the space of modular forms of weight k on Γ and by $\mathfrak{S}_k$ the space of *cusp forms*, defined by the additional requirement that $a_0 = 0$ in (45). The simplest and most famous examples of modular forms on Γ are the *Eisenstein series*

$$G_k(z) = -\frac{\mathbb{B}_k}{2k} + \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n$$

for $k \geq 4$ (here $\mathbb{B}_k$ is the kth Bernoulli number and $\sigma_\nu(n)$, as earlier in this paper, denotes the sum of the νth powers of the divisors of an integer n) and the *discriminant function* $\Delta(z)$ given by the Fourier expansion (17), which belong to $\mathfrak{M}_k$ and to $\mathfrak{S}_{12}$, respectively. The function $G_2(z)$ is not a modular form, but is related to Δ by $-48\pi i G_2(z) = \Delta'(z)/\Delta(z)$ and is "nearly modular" of weight 2 (more precisely, it satisfies $(G_2|_2\gamma)(z) = G_2(z) + \frac{i}{4\pi}\frac{c}{cz+d}$ instead of (44)). The space $\mathfrak{M}_k$ is finite-dimensional, with dimension given by

$$\dim \mathfrak{M}_k = n_k := \begin{cases} [k/12] & \text{if } k \equiv 2 \pmod{12}, \\ [k/12]+1 & \text{otherwise.} \end{cases}$$

This implies relations like $G_8 = 120G_4^2$ and $\Delta = 8000G_4^3 - 147G_6^2$. The space $\mathfrak{M}_k$ splits as $\langle G_k \rangle \oplus \mathfrak{S}_k$ for all $\kappa \geq 4$. It is 1-dimensional (spanned by G_k) precisely for five weights $k = 4$, 6, 8, 10 and 14, while $\mathfrak{M}_{12}$ is spanned by G_{12} and Δ.

On the other hand, an elementary calculation shows that $\mathfrak{W}_n$ is 1-dimensional, spanned by $x^n - 1$, for $n = 2$, 4, 6, 8 and 12, and is 3-dimensional, spanned by the three polynomials

$$x^{10} - 1, \quad x^8 - 3x^6 + 3x^4 - x^2, \quad 4x^9 - 25x^7 + 42x^5 - 25x^3 + 4x, \tag{46}$$

for $n = 10$, and more generally, that

$$\dim \mathfrak{W}_{k-2}^+ = n_k, \quad \dim \mathfrak{W}_{k-2}^- = n_k - 1$$

for all integers $k \geq 4$. This suggests that there might be canonical isomorphisms

$$r^+ : \mathfrak{M}_k \xrightarrow{\sim} \mathfrak{W}_{k-2}^+, \quad r^- : \mathfrak{S}_k \xrightarrow{\sim} \mathfrak{W}_{k-2}^-.$$

Such isomorphisms are given by the *theory of periods*. Let f be a form in $\mathfrak{M}_k$, with Fourier expansion given by (45). We define the *Eichler integral* of f by

$$\widetilde{f}(z) = \sum_{n=1}^{\infty} \frac{a(n)}{n^{k-1}} e^{2\pi i n z}$$

and set

$$r_f(z) = \widetilde{f}(z) - z^{k-2}\widetilde{f}(-1/z).$$

If f is a cusp form, then the $(k-1)$st derivative of $\widetilde{f}$ is a multiple of f, and from this and the modularity of f it follows by a simple calculation that the $(k-1)$st derivative of $r_f(z)$ is 0 and hence that $r_f \in \mathfrak{V}_{k-2}$. Equation (43) with $P = r_f$, $F = \widetilde{f}$ now implies that the polynomial $r_f(z)$ belongs to $\mathfrak{W}_{k-2}$, and the maps $r^{\pm} : \mathfrak{S}_k \to \mathfrak{W}_{k-2}^{\pm}$ are defined by sending f to the even or odd part of the polynomial $r_f(z)$. For f not cuspidal the situation is slightly different because the $(k-1)$st derivative of $\widetilde{f}$ differs from f by a constant, and hence r_f as defined above is no longer a polynomial. However, a simple calculation shows that it still is a linear combination of powers z^d with $-1 \leq d \leq k-1$, so that its even part is still a polynomial belonging to $\mathfrak{V}_{k-2}$ and hence to $\mathfrak{W}_{k-2}$, and in this way the map r^+ extends to all of $\mathfrak{M}_k$. It sends the Eisenstein series G_k to a multiple of the polynomial $X^{k-2} - 1 \in \mathfrak{W}_{k-2}^+$.

12. Eichler integrals on the real line, and generalized Dedekind symbols

As an example of what we just explained, take the cusp form $\Delta \in \mathfrak{S}_{12}$. We have

$$\widetilde{\Delta}(z) = \sum_{n=1}^{\infty} \frac{\tau(n)}{n^{11}} e^{2\pi i n z}, \quad \widetilde{\Delta}(z) - z^{10}\widetilde{\Delta}\Big(-\frac{1}{z}\Big) = r_\Delta(z) \quad (z \in \mathfrak{H}),$$

where r_Δ belongs to $\mathfrak{W}_{10}$ and hence is a linear combination of the three functions (46). In fact it has the form

$$r_\Delta(z) = \omega_+\Big(z^{10} - 1 - \frac{691}{36} z^2(z^2-1)^3\Big) + i\omega_- \frac{z(z^2-1)^2(4z^2-1)(z^2-4)}{1680}$$

for two real numbers ω_+ and ω_- (called the *periods* of Δ). On the other hand, it is known that $\tau(n) = \mathrm{O}(n^6)$ (or indeed $\mathrm{O}(n^c)$ for any $c > 11/2$, by Deligne's theorem), so the series defining $\widetilde{\Delta}$ converges on the real axis and we can write its value there as $\omega_+\Phi_+(x) + i\omega_-\Phi_-(x)$ with

$$\Phi_+(x) = \frac{1}{\omega_+}\sum_{n=1}^{\infty} \frac{\tau(n)}{n^{11}} \cos(2\pi n x), \quad \Phi_-(x) = \frac{1}{\omega_-}\sum_{n=1}^{\infty} \frac{\tau(n)}{n^{11}} \sin(2\pi n x). \tag{47}$$

Then $\Phi_+(x)$ is a continuous (indeed, 4 times differentiable) even function on $\mathbb{R}$ satisfying

$$\begin{aligned} \Phi_+(x+1) - \Phi_+(x) &= 0, \\ x^{10}\Phi_+\Big(\frac{1}{x}\Big) - \Phi_+(x) &= x^{10} - 1 - \frac{691}{36}x^2(x^2-1)^3 \end{aligned} \tag{48}$$

and $\Phi_-(x)$ is an odd continuous function satisfying similar transformation laws. Note that the periods $\omega_\pm$ are only canonically normalized up to rational factors. We have normalized ω_+ in such a way that $\Phi_+(0) = 1$. But then equations (48) imply by induction on the size of the numerator and denominator that $\Phi_+(x)$ is rational for every rational value of x, with a denominator dividing the 10th power of the denominator of x. In other words, the function on $\mathbb{Z}^2$ defined by

$$D_+(p,q) = \frac{q^{10}}{\omega_+}\sum_{n=1}^{\infty} \frac{\tau(n)}{n^{11}} \cos\frac{2\pi n p}{q} \tag{49}$$

takes on integer values. This function satisfies the homogeneity, periodicity and parity properties

$$D_+(kp,kq) = k^{10}D_+(p,q), \quad D_+(p+q,q) = D_+(p,q) = D_+(-p,q), \tag{50}$$

and the reciprocity law

$$D_+(p,q) - D_+(q,p) = q^{10} - p^{10} - \frac{691}{36}p^2q^2(q^2-p^2)^3, \tag{51}$$

and is uniquely characterized by these properties (and the normalization $D(1,0) = 0$). Similar properties apply to the function $D_-(p,q) = q^{10}\Phi_-(p/q)$, except that it is an odd function of p and satisfies the reciprocity law

$$D_{-}(p,q)+D_{-}(q,p)=\frac{1}{1680}pq(q^2-p^2)^2(p^2-4q^2)(4p^2-q^2). \qquad (52)$$

(The normalization of ω_{-} was chosen so that this function, too, takes on integer values with no common factor.) We call the functions $\Delta_{\pm}(p,q)$ the even and odd *Dedekind symbols* associated to the modular form Δ, because they are very analogous to the classical Dedekind sum or Dedekind symbol (which is also periodic in p and satisfies a reciprocity law like (52)). Actually, this is more than an analogy: the construction described here for Δ, which obviously can be carried through for any other cusp form on $SL(2,\mathbb{Z})$, also can be made sense of for Eisenstein series, including the "nearly modular" Eisenstein series G_2, and the odd Dedekind symbol attached to G_2 is precisely the classical Dedekind sum. (The odd Dedekind symbols attached to Eisenstein series of higher weights are classical analogues of the Dedekind sum involving higher Bernoulli polynomials rather than $\mathbb{B}_1$, while the even Dedekind symbol of G_k is the uninteresting function $(p,q)\mapsto q^{k-2}$.) We do not carry this out here, since the details are a bit technical and we have strayed somewhat from the main theme of this paper. We do, however, give a small table of values of the two Dedekind symbols associated to Δ (restricting to p and q coprime and $0\le p\le q/2$, as we can do by virtue of the homogeneity, periodicity, and parity properties):

q	p	$D_+(p,q)$	$D_-(p,q)$
1	0	1	0
2	1	−1049	0
3	1	−29399	20
4	1	12076	405
5	1	3132025	3564
	2	−8012423	2268
6	1	30839551	20020

q	p	$D_+(p,q)$	$D_-(p,q)$
7	1	178460401	84240
	2	−60349199	106920
	3	−259357199	48600
8	1	766572976	289170
	3	−765788624	298890
9	1	2690752401	852720
	2	647603601	1321320
	4	−3345823599	473640

The relationship between Dedekind sums and the period functions and Eichler integrals of modular forms has been observed by several people. It is discussed in detail in [Fu].

13. The modular explanation of $F_{k,D}(x)$

In terms of the notation just introduced, the assertion made in §5 about the function $\Phi_0(x)=\Phi(x)-\frac{360}{691}$ is simply that it is a multiple of the function $\Phi_+(x)$ defined by (47), the multiple being determined by any of the values of $\Phi(x)$ given in §4 to be $-\lambda = -\frac{360}{691}$. The proof is by now evident: the proof of Theorem 3 given in §6 shows that $\Phi_0(x)$ satisfies the same functional equations as in (48) except with the polynomials on the right multiplied by $-\lambda$, and these functional equations together with the continuity obviously determine the function uniquely. However, this is an ad hoc explanation and not very satisfactory, and one can ask whether the second term in (16) has a natural interpretation as the even part of the Eichler integral of some canonical cusp form of weight 12 associated to the discriminant D. In fact such a form exists, and is very simple.

To any $k > 2$ and any positive discriminant D we associate the function

$$f_{k,D}(z) = C_k D^{k-1/2} \sum_{b^2-4ac=D} \frac{1}{(az^2+bz+c)^k} \quad (z \in \mathfrak{H}), \tag{53}$$

where the sum is over *all* quadratic functions with integer coefficients and discriminant D and C_k is an unimportant normalizing factor. It is easily seen that $f_{k,D}$ is a cusp form on $SL(2,\mathbb{Z})$ of weight $2k$. These functions were introduced in the appendix of [Z1] for no particular reason other than their natural definition and analogy to Eisenstein series (in which one sums the $-k$th powers of all linear functions rather than of all quadratic functions of given discriminant). In [KZ1] and [Ko2] they were shown to have a more intrinsic meaning in terms of the theory of modular forms of half-integral weight (we will discuss this in more detail in §14), and in [KZ2] their even period functions were shown to be given by the formula

$$r^+(f_{k,D})(x) = \frac{\zeta_D(1-k)}{2\zeta(1-2k)}(X^{2k-1}-1) - P_{k,D}(x) \tag{54}$$

for all D (even including square D, with the definition of the polynomial $P_{k,D}(x)$ modified as in (30)). In view of (26), this means that the function $F_{k,D}(x)$ on $\mathbb{R}$ (defined by (15) for non-square D and by (29) when D is a square) has the Fourier expansion

$$F_{k,D}(x) = \frac{\zeta_D(1-k)}{2\zeta(1-2k)} + \sum_{n=0}^{\infty} \frac{a_{k,D}(n)}{n^{2k-1}} \cos(2\pi nx), \tag{55}$$

where $a_{k,D}(n)$ is the nth Fourier coefficient of the cusp form $f_{k,D}$.

The proof of (54), apart from (rather a lot of) technical details, is very simple. One of the standard formulas for the period polynomial $r_f(x)$ of a cusp form $f \in \mathfrak{S}_{2k}$ is as a multiple of the integral $\int_0^\infty f(z)(z-x)^{2k-2}dz$ (taken along the positive imaginary axis). If we apply this definition to $f = f_{k,D}$, look at the integrals coming from each summand separately (which we are not quite allowed to do; this is one of the technical details), symmetrize with respect to $x \mapsto -x$ (since we want only the *even* period polynomial) and combine the integrals coming from the quadratic functions $[a,b,c]$ and $[a,-b,c]$, then we obtain a sum of terms of the form

$$\int_{-\infty}^{\infty} \frac{(z-x)^{2k-2}}{(az^2+bz+c)^k} dz,$$

where the integral is now over the whole imaginary axis. But this integral is zero by the calculus of residues unless the zeros of the quadratic function ax^2+bx+c lie on opposite sides of the imaginary axis, and therefore the (main contribution to the) even period polynomial of $f_{k,D}$ is given by a finite sum over quadratic polynomials $[a,b,c]$ with discriminant D and $a>0>c$, as in (25). It is also possible to prove (55) directly, without considering (54) or making use of the results of [KZ2]: the computation of the constant Fourier coefficient of the periodic function $F_{k,D}(x)$ was given in §8, and a rather similar calculation permits one to write the higher Fourier coefficients in a form which can be compared conveniently with the Fourier expansion of $f_{k,D}(z)$. The details are left to the interested reader.

All of these modular calculations carry over to the case of equivalence classes of quadratic forms which was discussed in §9. For each such class $\mathcal{A}$ one defines a cusp form $f_{k,\mathcal{A}}(z)\in\mathfrak{S}_{2k}$ by restricting the summation in (53) to $[a,b,c]\in\mathcal{A}$, and then by taking appropriate symmetrizations and antisymmetrizations one can get both the odd and even parts of $P^*_{k,\mathcal{A}}(x)$ (resp. $F^*_{k,\mathcal{A}}(x)$) as odd or even parts of the period polynomials (resp. Eichler integrals restricted to the real line) of these cusp forms. We refer the reader to [KZ2] for the details.

14. The relation to modular forms of half-integral weight

Modular forms of half-integral weight have been known for a long time, the simplest example being the Jacobi theta function $\theta(z)=\sum_{n=-\infty}^{\infty}q^{n^2}$ which transforms under the action of the group $\Gamma_0(4)=\bigl\{\bigl(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\bigr)\mid c\equiv 0\pmod 4\bigr\}$ by $\theta\bigl(\frac{az+b}{cz+d}\bigr)=\varepsilon(cz+d)^{1/2}\theta(z)$ for some 4th root of unity ε. They became of great interest in the 70's because of Shimura's discovery of a correspondence between modular forms of half-integral and integral weights. In general this correspondence is quite complicated and is neither surjective nor injective, but in certain simple cases Kohnen [Ko1] found versions of the Shimura lifting which are isomorphisms. In particular, if we define $\mathfrak{M}_{k+1/2}$ for any even integer $k\ge 0$ to be the space of functions $g(z)$ such that $g(z)/\theta(z)$ transforms under $\Gamma_0(4)$ like a modular form of weight k and g has a Fourier expansion of the form

$$g(z)=\sum_{D\ge 0,\ D\equiv 0\text{ or }1\pmod 4}c(D)q^D\quad(q=e^{2\pi iz})\tag{56}$$

("Kohnen plus-space"), then $\mathfrak{M}_{k+1/2}$ and $\mathfrak{M}_{2k}$ are isomorphic as vector spaces, and even as modules over the ring of Hecke operators. This latter point says that for modular forms $f\in\mathfrak{M}_{2k}$ having multiplicative Fourier coefficients, like G_k or Δ, there is a corresponding form $g\in\mathfrak{S}_{k+1/2}$, unique up to multiplication by a scalar, such that the quotient of $c(Dp^2)$ by $c(D)$ is given by a simple formula involving the pth Fourier coefficient of f. In particular, the form corresponding to G_k is the modular form $\mathcal{H}_k(z)=\sum_{D\ge 0}H(k,D)q^D$ mentioned in §8, and the Hecke compatibility property is equation (31). For $k=2$ and $k=4$ this says that the series

$$\mathcal{A}(z)=\sum_{D\ge 0}\alpha_Dq^D=-\tfrac{1}{24}+\tfrac{5}{12}q+\tfrac{35}{12}q+2q^5+5q^8+\tfrac{125}{12}q^9+10q^{12}+10q^{13}+\dots$$

and

$$\mathcal{B}(z) = \sum_{D\ge 0} \beta_D q^D = \tfrac{1}{240} + \tfrac{1}{120}q + \tfrac{121}{120}q^4 + 2q^5 + 11q^8 + \tfrac{2161}{120}q^9 + 46q^{12} + 58q^{13} + \dots$$

(where we have taken the coefficients from the tables in Sections 2, 3 and 7) belong to $\mathfrak{M}_{5/2}$ and $\mathfrak{M}_{9/2}$, respectively. For the latter function this is immediately obvious from the formula (10) for β_D (to which $1/120$ must be added when D is a square because of the second term in (30)), since this equation just says that $\mathcal{B}(z)$ is the product of $\theta(z) = 1+2q+2q^4+2q^9+\dots$ with $G_4(4z) = \frac{1}{240}+q^4+9q^8+28q^{12}+\dots$, which is a modular form of weight 4 on $\Gamma_0(4)$ because $G_4(z)$ is a modular form of weight 4 on $SL(2,\mathbb{Z})$. For $\mathcal{A}(z)$ the idea is the same but this time $\mathcal{A}(z)$ is given by

$$\mathcal{A}(z) = \theta(z)G_2(4z) + \frac{1}{2}\sum_{m=1}^{\infty} m^2 q^{m^2} = \theta(z)G_2(4z) + \frac{1}{8\pi i}\theta'(z),$$

where the second term comes from the second correction term in the definition of α_{m^2} given in §7, and the modularity follows from the "near modularity" of G_2 mentioned in §11.

These two examples would make us expect that also the coefficients γ_D and δ_D in (14) are the coefficients of a modular form of half-integral weight, and indeed this is true. In particular, the cusp form in $\mathfrak{M}_{13/2}$ corresponding to the discriminant function $\Delta \in \mathfrak{M}_{12}$ is the function

$$\mathcal{D}(z) = \sum_{D>0} \delta_D q^D = \tfrac{1}{120}q - \tfrac{7}{15}q^4 + q^5 - 2q^8 + \tfrac{3}{40}q^9 + 12q^{12} - 11q^{13} + \dots.$$

This can be seen "by hand" by observing that $10\delta_D$ equals the coefficient γ'_D in (24), i.e.

$$\delta_D = \frac{1}{10}\gamma'_D = \frac{1}{10}\sum_{\substack{b^2-4ac=D\\ a>0>c}} (10a^3b^2 + 5a^4c) = \sum_{\substack{|b|<\sqrt{D}\\ b\equiv D \pmod 2}} \frac{9b^2-D}{8}\sigma_3\Big(\frac{D-b^2}{4}\Big),$$

where if D is a square the quantity $\sigma_3(0)$ which then appears in the formula must be interpreted as $1/240$, the constant term of the Eisenstein series G_4. This gives the representation

$$\mathcal{D}(z) = \frac{1}{2\pi i}G_4(4z)\theta'(z) - \frac{1}{8\pi i}G'_4(4z)\theta(z)$$

of $\mathcal{D}(z)$, and the modularity follows. The same works for the other coefficients; for instance, one can give a modular proof of the vanishing of the coefficient β'_D in (22) (which we deduced in §6 in an elementary way from the "period polynomial" property of $P_{4,D}$) by writing

$$\beta'_D = \sum_{\substack{b^2-4ac=D\\ a>0>c}} (3ab^2 + 3a^2c) = 3\sum_{\substack{|b|<\sqrt{D}\\ b\equiv D \pmod 2}} \frac{5b^2-D}{4}\sigma_1\Big(\frac{D-b^2}{4}\Big)$$

and deducing a representation of $\sum \beta'_D q^D$ as an element of $\mathfrak{M}_{9/2}$ with constant term 0 and hence as the zero form. The method works in general and expresses the coefficient of x^{2n} in $P_{k,D}(x)$ for every n as the Dth Fourier coefficient of a modular form of weight $k+1/2$ which is a combination of derivatives (the so-called "Cohen bracket", introduced in [Co3]) of an Eisenstein series and the function θ. (More precisely, for $n < k/2$ it is the nth Cohen bracket of $G_{k-2n}(4z)$ and $\theta(z)$.) The details of the calculation, which are quite easy, are given in [KZ2], pp. 218–219. The modularity property in question also follows from equation (55), since it was shown in [KZ1] and [KZ2] that the function $\Omega_k(z,z') = \sum_D f_{k,D}(z)e^{2\pi i D z'}$ is a modular form of weight $k+\frac{1}{2}$ in z' and is in fact the "kernel function" for the above-mentioned Shimura correspondence between the spaces $\mathfrak{S}_{k+1/2}$ and $\mathfrak{S}_{2k}$. (This means that the Petersson scalar product of any function $f \in \mathfrak{S}_{2k}$ with $\Omega_k(\cdot, z')$ is the Shimura lift of f.)

In summary, we have indicated in this section the proof, or several proofs, of the fact that the function $T_x(z) := \sum_{D\geq 0} F_{k,D}(x)q^D$ is a modular form of weight $k+\frac{1}{2}$ for any real number x. Substituting the definition of $F_{k,D}$ from (15), (29) (resp. (28) in the case $k=2$) and (33), we find that $T_x(z)$ has the expansion

$$\sum_{\substack{(a,b,c)\in\mathbb{Z}^3\\ ax^2+bx+c>0>a}} (ax^2+bx+c)^{k-1}q^{b^2-4ac} \\ - \frac{1}{2k}\sum_{m=-\infty}^{\infty} \bar{\mathbb{B}}_k(mx)q^{m^2} + \delta_{k,2}\frac{\kappa(x)}{2}\sum_{m=1}^{\infty} m^2 q^{m^2}. \tag{57}$$

(Note that, although we obtained this by expanding the double sum

$$\sum_{D\geq 0}\sum_{b^2-4ac=D},$$

we do not have to write the condition $b^2-4ac>0$ explicitly in the first summation in (57) because it follows automatically from the inequalities $a<0<ax^2+bx+c$ by the same identity $b^2-4ac=(2ax+b)^2-4a(ax^2+bx+c)$ which played a role in §1.) This formula is interesting because it has the form of a general theta series, i.e., a sum of the form $\sum P(x)q^{Q(x)}$ where Q is a quadratic form on a lattice L and P a spherical function with respect to the quadratic form Q. In the classical case when Q is a positive definite quadratic form and P a homogeneous polynomial, it is known that such a theta series is a modular form of weight equal to the sum of the degree of P and half the rank of L. Here the sum has the same form (except for the two last sums in (57), which are much "thinner" than the first one and must be interpreted as some kind of a boundary term), except that the quadratic function $(a,b,c)\mapsto b^2-4ac$ is indefinite and the summation is only over a cone (defined by the two linear inequalities $a<0$ and $ax^2+bx+c>0$, x being fixed), but the final result is still the same. This example suggests that there may be an arithmetic theory of theta series attached to indefinite quadratic forms in which the summation runs over the intersection of the lattice with some simplicial cone on which the quadratic form is positive and the result is still a modular form of the expected weight and level.

References

[ChZ] Choie, Y-J., Zagier, D., Rational period functions for $PSL(2,\mathbb{Z})$. In: A tribute to Emil Grosswald: number theory and related analysis (ed. by M. Knopp and M. Sheingorn; Contemp. Math. 143), 89–108. American Mathematical Society, Providence 1993.

[Co1] Cohen, H., Sommes de carrés, fonctions L et formes modulaires. C. R. Acad. Sci. Paris Sér. A Sci. Math. 277 (1973), 827–830.

[Co2] — Variations sur un thème de Siegel et Hecke. Acta Arith. 30 (1976), 63–93.

[Co3] — Sums involving the values at negative integers of L-functions of quadratic characters. Math. Ann. 217 (1975), 271–285.

[Fu] Fukuhara, S., Modular forms, generalized Dedekind symbols and period polynomials. Math. Ann. 310 (1998), 83–101.

[Ko1] Kohnen, W., Modular forms of half-integral weight on $\Gamma_0(4)$. Math. Ann. 248 (1980), 249–266.

[Ko2] — Beziehungen zwischen Modulformen halbganzen Gewichts und Modulformen ganzen Gewichts (Bonner Math. Schriften 131). Univ. Bonn, 1981.

[KZ1] Kohnen, W., Zagier, D., Values of L-series of modular forms at the center of the critical strip. Invent. Math. 64 (1980), 175–198.

[KZ2] — Modular functions with rational periods. In: Modular forms (ed. by R.A. Rankin), 197–249. Ellis Horwood, Chichester 1984.

[Sh] Shimura, G., On modular forms of half-integral weight. Ann. of Math. (2) 97 (1973), 440–481.

[Si] Siegel, C.L., Berechnung von Zetafunktionen an ganzzahligen Stellen. Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II 10 (1969), 87–102.

[Z1] Zagier, D., Modular forms associated to real quadratic fields. Invent. Math. 30 (1975), 1–46.

[Z2] — On the values at negative integers of the zeta-function of a real quadratic field. Enseign. Math. (2) 22 (1976), 55–95.

[Z3] — Modular forms whose Fourier coefficients involve zeta-functions of quadratic fields. In: Modular functions of one variable, VI (ed. by J.-P. Serre and D.B. Zagier; Lecture Notes in Math. 627), 105–169. Springer, Berlin 1977.

[Z4] — Quelques conséquences surprenantes de la cohomologie de $SL(2,\mathbb{Z})$. In: Leçons Mathématiques de Bordeaux. Editions Cassini (to appear).