

# RIEMANN SURFACES AND NUMBER THEORY

BROOKS ROBERTS

ABSTRACT. The national bestseller *Fermat's Enigma* describes modular forms as “some of the most bizarre and wonderful objects in mathematics”. In this talk we will look at some examples of how modular forms arise in the number theory of quadratic forms. The construction will involve a number of interesting topics, including Fourier analysis, group actions, and compact Riemann surfaces.

A basic problem in number theory is to say something about the solutions of diophantine equations. Suppose we are given a polynomial  $p(x, y, z, \dots)$  in several variables with integer coefficients, and another integer  $n$ . Then what can we say about the solutions to

$$p(x, y, z, \dots) = n$$

in the integers? For example, how many integral solutions are there? One can try to attack the problem by looking at classes of polynomials. The case of linear polynomials is handled by the Euclidean algorithm. In this talk we will look at the case of quadratic polynomials. Let  $Q(x_1, \dots, x_m)$  be a quadratic form, i.e., a polynomial of the form

$$Q(x) = \sum_{i,j=1}^m a_{ij}x_i x_j,$$

where

$$A = (a_{ij})$$

is a symmetric matrix with integer coefficients. A simple example is

$$Q(x) = x_1^2 + \dots + x_m^2,$$

so that our question becomes in how many ways can a positive integer be represented as a sum of squares? To make the exposition simpler, we will consider another class of examples. We assume that

- $Q(x)$  is positive definite, i.e,  $Q(x) > 0$  for real nonzero  $x$ ,
- $\det(A) = 1$ ,
- $Q(x)$  is even for integral  $x$ .

The last condition just means the diagonal entries of  $A$  are even. These conditions are rather restrictive. For example, the number of variables  $m$  must be divisible by 8, and the number of such quadratic forms for each  $m$  is finite. Here is a table for the number of inequivalent such quadratic forms for the first few values of  $m$ :

$m$	Number of $Q(x)$
8	1
16	2
24	24
32	> 80 million

If  $m = 8$  then the single such quadratic form is given by

$$A = \begin{pmatrix} 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}$$

This quadratic form comes from the exceptional irreducible root system  $E_8$ . If  $m = 24$ , then one of the  $Q(x)$  comes from the Leech lattice; its integral automorphism group (i.e., the group of  $m$  by  $m$  integral nonsingular matrices  $B$  such that  ${}^tBAB = A$ ) has order 8,315,553,613,086,720,000. The quotient of this automorphism group by  $\pm 1$  is one of the finite number of sporadic simple groups.

Turning back to our original problem suppose we are given a quadratic form  $Q(x)$  satisfying the above three conditions. Given an integer  $n$ , we want to say something about the integral solutions to

$$Q(x) = n.$$

Obviously, if  $n$  is less than zero or odd there are no solutions. So let us assume that  $n$  is a nonnegative integer and let

$$a(n) = \text{number of integral solutions } x \text{ to } Q(x) = 2n.$$

We want to say something about the sequence

$$a(0), a(1), a(2), \dots$$

Evidently, we have  $a(0) = 1$ , with  $x = 0$  the only solution to  $Q(x) = 0$ . In this talk we will give a formula for  $a(n)$  when the number of variables is  $m = 8$  or 16. We will show that for  $n \geq 1$ ,

$$a(n) = \begin{cases} 240 \times \text{sum of the cubes of the positive divisors of } n & \text{if } m = 8 \\ 480 \times \text{sum of the seventh powers of the positive divisors of } n & \text{if } m = 16. \end{cases}$$

Note that this implies that the two distinct quadratic forms for  $m = 16$  have the *same* number of solutions! For a general number of variables our method also has something to say about  $a(n)$  but the result is more of a qualitative nature.

Our first step towards proving these results will be to introduce a certain analytic function built up from the  $a(n)$ ; after establishing the essential properties of this function we will explain how it will help us understand the  $a(n)$ . We begin with a naive idea. Recall that when dealing with a sequence of numbers  $a(n)$  it is often useful to define their formal generating function:

$$f(q) = a(0) + a(1)q + a(2)q^2 + \dots$$

where  $q$  is a formal variable. Now if, for example, our sequence satisfied some linear recurrence relation, then we could solve for the  $a(n)$  via some simple algebraic manipulations. However, our case is more complicated, and we will need to regard the generating function as an analytic function and then introduce methods from analysis and geometry. Let us determine the convergence of  $f(q)$  when  $q$  is a complex variable. The power series clearly diverges if  $q = 1$ . However we claim that the series converges in the unit disk of complex numbers  $q$  with  $|q| < 1$ . We have formally

$$f(q) = a(0) + a(1)q + a(2)q^2 + \dots = \sum_{x \in \mathbb{Z}^m} q^{Q(x)/2}.$$

Since  $Q(x)$  is a positive definite quadratic form,

$$\sqrt{Q(x)}$$

defines a norm on  $\mathbb{R}^m$ . All norms on  $\mathbb{R}^m$  are equivalent, so there exists a positive constant  $c$  such that

$$Q(x) \geq c(x_1^2 + \dots + x_m^2), \quad x \in \mathbb{R}^m.$$

Hence for  $|q| < 1$

$$\sum_{x \in \mathbb{Z}^m} |q^{Q(x)/2}| \leq \sum_{x \in \mathbb{Z}^m} |q|^{c(x_1^2 + \dots + x_m^2)/2} \leq \left( \sum_{x \in \mathbb{Z}} |q|^{cx^2/2} \right)^m.$$

This last series converges by comparison to the geometric series, proving our claim. Having established that our generating function is actually analytic in the unit disk, we will change things slightly by making a simple change of variables. We will compose  $f(q)$  with the function from the upper half plane  $\mathbb{H}^2$  of complex numbers with positive imaginary part to the unit disk given by

$$\mathbb{H}^2 \rightarrow D, \quad z \mapsto q = e^{2\pi iz} = e^{2\pi ix} e^{-2\pi y}.$$

The significance of the upper half plane will be made clear later in the talk. We note that this change of variables is not onto as it misses 0. Also, it is not one-to-one as any two

vertical lines which are one unit apart get mapped to the same line in the disk. We now set

$$\theta(z) = f(q), \quad z \in \mathbb{H}^2.$$

The analytic function  $\theta(z)$  will be our basic object of study. It is an example of a **theta series**.

The key property of  $\theta(z)$ , which we will now establish, is that it satisfies two functional equations. The first equation is

$$\theta(z + 1) = \theta(z), \quad z \in \mathbb{H}^2.$$

This is obvious. However,  $\theta(z)$  also satisfies a nontrivial symmetry which is not obvious but not too difficult to prove. We claim that

$$\theta(-1/z)(-z)^{-m/2} = \theta(z), \quad z \in \mathbb{H}^2.$$

To prove this, we note first that since both sides are analytic functions in the upper half plane, by the identity principle it suffices to show this for  $z = it$  on the vertical line through  $i$ . For  $t > 0$  we have

$$\theta(z) = \theta(it) = \sum_{x \in \mathbb{Z}^m} e^{-\pi t Q(x)} = \sum_{x \in \mathbb{Z}^m} g_t(x),$$

where

$$g_t(x) = e^{-\pi t Q(x)}, \quad x \in \mathbb{R}^m.$$

Now an important connection between the continuous and the discrete is made by the **Poisson summation formula** which asserts that for any sufficiently rapidly decaying function (i.e., Schwarz function)  $g$  on  $\mathbb{R}^m$ ,

$$\sum_{x \in \mathbb{Z}^m} g(x) = \sum_{x \in \mathbb{Z}^m} \hat{g}(x),$$

where  $\hat{g}$  is the Fourier transform of  $g$ :

$$\hat{g}(x) = \int_{\mathbb{R}^m} g(y) e^{-2\pi i x \cdot y} dy.$$

To compute  $\theta(it)$  it would thus suffice to compute  $\hat{g}_t(x)$ . This is easy to do if we first diagonalize  $A$ , and then use the fact that the Fourier transform in one variable of the Gaussian

$$e^{-\pi x^2}$$

is itself. The result is that  $\theta(z)$  satisfies the above functional equation.

So far, we have introduced an analytic function constructed from the sequence  $a(n)$  and shown that it satisfies some curious functional equations. But what really is this strange function  $\theta(z)$ , and can it help us understand the  $a(n)$ ? The strategy for the remainder of this talk will be to look not just at  $\theta(z)$ , but at the set of all analytic functions  $h(z)$  on the upper half plane which have the same properties as  $\theta(z)$ , i.e., which have an absolutely convergent expansion

$$h(z) = \sum_{n=0}^{\infty} b(n)q^n, \quad z \in \mathbb{H}^2,$$

and which satisfy the same functional equations as  $\theta(z)$ . These functions are called **modular forms** (for the full modular group), and the set of all modular forms (for the full modular group) is clearly a complex vector space. Though it is not at all obvious, the functional equations are quite restrictive: We will show in the remainder of the talk that the vector space of modular forms has dimension

$$1 + [m/24],$$

which is one when  $m = 8$  or  $16$ ! In doing so, we will relate modular forms to noneuclidean geometry and compact Riemann surfaces, which were extensively studied in the last century by Klein, Poincare and others, and so modular forms won't seem so curious after all. Finally, from the one dimensionality of the space of modular forms when  $m = 8$  or  $16$  we will deduce the promised formulas for  $a(n)$ .

We first need to reformulate the functional equations in terms of group theory. We begin by observing that the two symmetries of upper half plane  $\mathbb{H}^2$

$$z \mapsto z + 1, \quad z \mapsto -1/z$$

in the functional equations are examples of **fractional linear transformations**. If

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is nonsingular matrix with complex entries, then the function

$$z \mapsto \alpha \cdot z = \frac{az + b}{cz + d}$$

defines an automorphism of the extended complex plane called a fractional linear transformation. Evidently, we have

$$z + 1 = T \cdot z, \quad -1/z = S \cdot z$$

where

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

As linear fractional transformations,  $T$  and  $S$  map the upper half plane onto itself. Now it is a basic fact that if  $\alpha$  and  $\beta$  are nonsingular two by two complex matrices, then

$$(\alpha\beta) \cdot z = \alpha \cdot (\beta \cdot z)$$

for  $z$  in the extended complex plane; in other words, the group of such matrices acts on the extended complex plane. From this, we deduce that the group of matrices generated by  $T$  and  $S$  acts on the upper half plane. As it turns out, this group is simply the group of all two by two integral matrices of determinant one, i.e.,

$$\mathrm{Sl}(2, \mathbb{Z}).$$

We now have conceptual characterization of the symmetries in the functional equations: they are generators for the action of  $\mathrm{Sl}(2, \mathbb{Z})$  on the upper half plane.

How can we use the group action to help compute the dimension of the space of modular forms? To see how, we use a clever trick. Fix one nonzero modular form; we may as well take  $\theta(z)$ . Then the map

$$h(z) \mapsto (h/\theta)(z) = h(z)/\theta(z)$$

is a map from the space of modular forms to a certain space of functions on the upper half plane. Of course, each function  $h(z)/\theta(z)$  is not holomorphic, but it is meromorphic. Also, the map  $h(z) \mapsto (h/\theta)(z)$  is clearly an injective map, so that its image has the same dimension as the space of modular forms. Moreover, and this is the point of this map, since  $S$  and  $T$  generate  $\mathrm{Sl}(2, \mathbb{Z})$  a glance at the functional equations shows that  $(h/\theta)(z)$  is invariant under the action of  $\mathrm{Sl}(2, \mathbb{Z})$

$$(h/\theta)(\alpha \cdot z) = (h/\theta)(z), \quad \alpha \in \mathrm{Sl}(2, \mathbb{Z}), z \in \mathbb{H}^2.$$

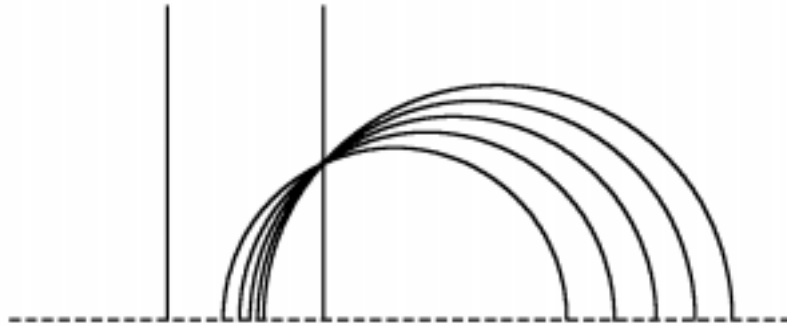
Thus, we may regard  $(h/\theta)(z)$  as a function on the quotient

$$\mathrm{Sl}(2, \mathbb{Z}) \backslash \mathbb{H}^2,$$

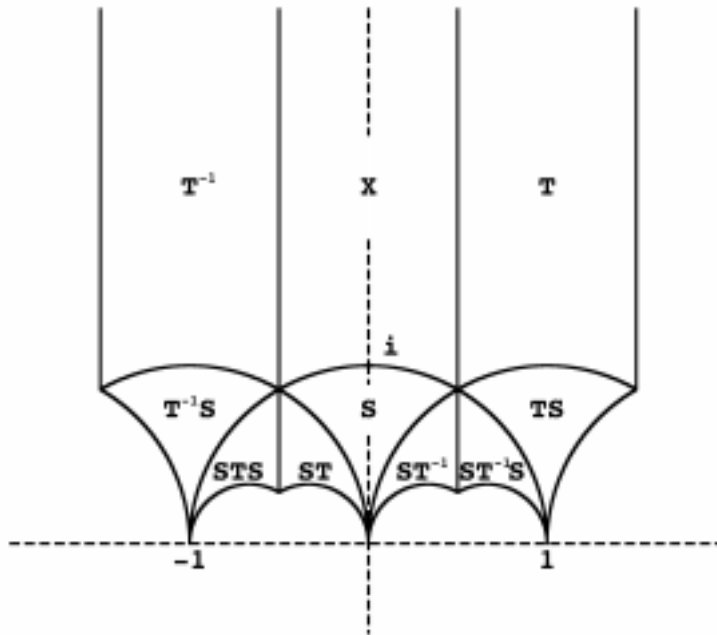
which may not be defined at all points. This translation of the problem is fruitful: As it turns out, the quotient  $\mathrm{Sl}(2, \mathbb{Z}) \backslash \mathbb{H}^2$  is just the Riemann sphere with one point removed, and the space of functions on the quotient that we just created turns out to be exactly the sort of space of meromorphic functions whose dimension can be computed via the Riemann-Roch theorem. To explain how this works we must first talk about the action of  $\mathrm{Sl}(2, \mathbb{Z})$  on the upper half plane.

To understand the action of  $\mathrm{Sl}(2, \mathbb{Z})$  it is useful to recall the noneuclidean nature of the upper half plane  $\mathbb{H}^2$ . Just as the Euclidean plane comes equipped with the Euclidean metric, which has constant sectional curvature 0 everywhere, the upper half plane admits a Riemannian metric with constant sectional curvature  $-1$  everywhere. Another name for the upper half plane with this metric is the hyperbolic plane: this explains the notation

$\mathbb{H}^2$ . One can easily give a formula for this metric, but more intuitively the geodesics, i.e., length minimizing lines, in  $\mathbb{H}^2$  are either vertical lines, or semi-circles abutting the real axis, as in the picture below:



The geodesics (lines) in the hyperbolic plane  $\mathbb{H}^2$  give a model for a noneuclidean geometry: for any line and point not on the line, there are infinitely many lines through the point parallel to the given line (i.e., not intersecting the given line). This is illustrated in one case in the above picture. The connection between the action of  $Sl(2, \mathbb{Z})$  and the hyperbolic structure on the upper half plane is that the elements of  $Sl(2, \mathbb{Z})$  act as isometries of the hyperbolic metric, and thus preserve geodesics. In fact, one can show that action of  $Sl(2, \mathbb{Z})$  is described by the following picture.



Here, each region (really a triangle) is obtained from the region labelled 1 via the specified element. Moreover, the interior of each region is a fundamental domain for the action of  $Sl(2, \mathbb{Z})$  (i.e, no two points in the interior of the given region are equivalent, and any point in the plane is equivalent to a point in the interior of a region or its boundary). We can be

more precise about how points on the boundaries are related via  $\mathrm{Sl}(2, \mathbb{Z})$ . Suppose that we just consider the region 1. Then the left vertical boundary is taken into the right vertical boundary by  $T$ , and the portion of the lower boundary to the left of  $i$  is taken into the portion to the right of  $i$  via  $S$ , with  $i$  being fixed. Also, no two distinct points on the left vertical boundary and no two points on the boundary arc to the left of  $i$  are equivalent. It follows that there is a homeomorphism

$$\mathrm{Sl}(2, \mathbb{Z}) \backslash \mathbb{H}^2 \cong X / \sim$$

where  $\sim$  is the identification of the left and right vertical boundaries of  $X$ , and the identification of the left and right horizontal boundaries of  $X$ .

Using this understanding, we can moreover make  $\mathrm{Sl}(2, \mathbb{Z}) \backslash \mathbb{H}^2$  into a Riemann surface, and by adding an ideal point, into a compact Riemann surface isomorphic to the Riemann sphere. To give the quotient a complex structure we must specify a small open set around each point  $p$  along with a coordinate map, i.e., a homeomorphism to an open disk centered at 0 in the complex plane; also, one must show that the change of coordinate maps are analytic. The easiest case is if  $p$  corresponds to a point  $z$  in the interior of  $X$ . In this case we let the open set containing  $p$  be any open set corresponding to a small open disk centered at  $z$  and contained in the interior of  $X$ ; such a disk is a translation of an open disk centered at 0, and so we obtain coordinate map. The cases when  $p$  corresponds to a boundary point of  $X$  are more challenging, and we will not describe them. Instead, let us concentrate on how to compactify  $\mathrm{Sl}(2, \mathbb{Z}) \backslash \mathbb{H}^2$  and make it into a compact Riemann surface. From the way the boundary points of  $X$  are identified, it is clear how to compactify  $\mathrm{Sl}(2, \mathbb{Z}) \backslash \mathbb{H}^2$  as a topological space: we add one point,  $\infty$ , and define a basis for the open sets containing  $\infty$  to be the union of  $\infty$  with the sets of points corresponding to points  $z$  in  $X$  such that  $|z| > C$  for some fixed  $C > 0$ . As a coordinate neighborhood for  $\infty$  along with a coordinate map we take one of these open sets, and define the coordinate map by

$$z \mapsto q, \quad z \neq \infty, \quad \infty \mapsto 0;$$

of course, this is the change of coordinates we introduced earlier, and we will see the significance of this definition in a moment. With these definitions, one can check that  $\mathrm{Sl}(2, \mathbb{Z}) \backslash \mathbb{H}^2 \cup \{\infty\}$  is a compact Riemann surface; clearly, it has genus 0 and is thus isomorphic to the Riemann sphere.

Now let us return to our problem of computing the dimension of the space of modular forms. We recall the map from the space of modular forms to a space of functions (possibly not defined at some points) on the quotient  $\mathrm{Sl}(2, \mathbb{Z}) \backslash \mathbb{H}^2$  given by

$$h(z) \mapsto (h/\theta)(z).$$

Using the definition of the complex structure on the quotient, it is not too difficult to see that  $(h/\theta)(z)$  gives a meromorphic function on  $\mathrm{Sl}(2, \mathbb{Z}) \backslash \mathbb{H}^2$ . Moreover, we see that

$(h/\theta)(z)$  extends to a meromorphic function on  $\mathrm{Sl}(2, \mathbb{Z}) \backslash \mathbb{H}^2 \cup \{\infty\}$  because for points  $p$  in our neighborhood of  $\infty$  corresponding to points  $z$  in  $X$  with  $|z| > C$  we have

$$(h/\theta)(z) = \sum_{n=0}^{\infty} b(n)q^n / \sum_{n=0}^{\infty} a(n)q^n.$$

The expansion condition in the definition of a modular form is thus exactly what is required to regard  $(h/\theta)(z)$  as being meromorphic at  $\infty$ ! To now compute the dimension of the space of modular forms one can use the Riemann-Roch theorem. One can show that there exists finitely many points  $p_1, \dots, p_l$  on  $\mathrm{Sl}(2, \mathbb{Z}) \backslash \mathbb{H}^2 \cup \{\infty\}$  and integers  $n_1, \dots, n_l$  such that the image of the map  $h(z) \mapsto (h/\theta)(z)$  is exactly the set of meromorphic functions  $F$  on  $\mathrm{Sl}(2, \mathbb{Z}) \backslash \mathbb{H}^2 \cup \{\infty\}$  which are holomorphic at all the points not among the  $p_i$ , and which have order larger than or equal to the  $n_i$  at  $p_i$ . The Riemann-Roch theorem asserts that if

$$-(n_1 + \dots + n_l) > 2g - 2,$$

where  $g$  is the genus of the compact Riemann surface (in our case  $g = 0$ ), then the dimension of the space of such meromorphic functions  $F$  is

$$-(n_1 + \dots + n_l) - g + 1.$$

In our case it turns out that

$$-(n_1 + \dots + n_l) = [m/24],$$

where  $[m/24]$  is the integer just to the left of  $m/24$ . Thus, the dimension of the space of modular forms is

$$1 + [m/24].$$

Finally, we deduce the promised formulas for  $a(n)$  when  $m = 8$  or  $16$ . When  $m = 8$  or  $16$ , we see that the dimension of the space of modular forms is 1, so that this space is spanned by  $\theta(z)$ . On the other hand, the space of modular forms always contains the following nonzero element

$$E(z) = \frac{1}{2\zeta(m/2)} \sum_{(a,b) \in \mathbb{Z}^2, (a,b) \neq (0,0)} \frac{1}{(az + b)^{m/2}}$$

which is called an **Eisenstein series**. It is easy to check that the Eisenstein series satisfies the two functional equations. One can show that the Eisenstein series also satisfies the expansion condition; in fact, using the formula

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

one finds that

$$E(z) = 1 + \frac{(2\pi i)^{m/2}}{2\zeta(m/2)(m/2 - 1)!} \sum_{n=1}^{\infty} \sigma_{m/2-1}(n)q^n,$$

where  $\sigma_{m/2-1}(n)$  is the sum of the  $m/2 - 1$  powers of the positive divisors of  $n$ . In particular, when  $m = 8$ , one has

$$E(z) = 1 + 240 \sum_{n=1}^{\infty} \sigma_3(n)q^n,$$

and if  $m = 16$ ,

$$E(z) = 1 + 480 \sum_{n=1}^{\infty} \sigma_7(n)q^n.$$

If  $m = 8$  or  $16$  we thus must have

$$\theta(z) = cE(z)$$

for some constant  $c$ . Using that we know  $a(0) = 1$  we get that  $c = 1$ . This gives the desired formulas for  $a(n)$  when  $m = 8$  or  $16$ .

BROOKS ROBERTS, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF IDAHO, MOSCOW, ID 83844  
USA

*E-mail address:* brooksr@uidaho.edu