Research Academy for Young Scientists

Modular Forms on the Moduli Space of Complex Elliptic Curves

Daniel Nordström
karl.daniel.nordstrom@gmail.com

under the direction of
Sjoerd Wijnand de Vries
Stockholm University
Department of Mathematics

July 12, 2023

*

Abstract

Modular forms, moduli spaces and elliptic curves are all fundamental concepts in modern abstract algebra. This paper aims to find the moduli space of complex elliptic curves and to study its properties. In particular, we show that the moduli space we find allows for defining modular forms in a natural and useful way. This is accomplished by using the fundamental domain for a certain matrix group as a parameterization for the elliptic curves. Along the way, we study the j-invariant, which classifies isomorphism classes of elliptic curves. The objects discussed are all subject to active research and have applications in areas from topology and geometry to applied number theory such as cryptography.

Acknowledgements

I want to express my sincere gratitude towards my mentor Sjoerd Wijnand de Vries for investing so much time into guiding me through this project, and for giving me a new perspective on research through continually coming up with new challenges for me to dive into. I would also like to thank Julia Mårtensson, Linnea Jonröd, Max Eriksson and Viktor Sundström for arranging Rays 2023, thus enabling me to undertake this project. Finally, I would like to thank Rays – for excellence and their partners Beijerstiftelsen and Kungliga Patriotiska Sällskapet.

Contents

1 Introduction

Modular forms and elliptic curves are widely studied concepts in modern mathematics since they exhibit remarkable properties, and since there is a profound connection between them. This link between elliptic curves and modular forms was first proven by Andrew Wiles; although he only demonstrated it for a significant special case, it was sufficient to imply the negative resolution of the long-open Fermat's Last Theorem, which states that the equation $x^n + y^n = z^n$ has no solution in positive integers for $n \geq 3$.

Formally, modular forms are a kind of complex-valued functions which are smooth. This means that their complex derivative exists everywhere that they are defined. They also have a value at infinity, which means that the limit of the function as its argument tends to infinity exists. Finally, they are periodic, and one can manipulate them and their arguments in certain ways without changing their values.

The set of solutions to an equation in two variables can be described by a so-called *graph*, a set of points in the plane. Higher-dimensional graphs exist for equations using more variables; for example, a plane can be seen as a graph in three-dimensional space. Elliptic curves are the graphs of certain polynomial equations of degree 3, on which we can define addition of two points so that we may carry out basic arithmetic on the curve itself. As it turns out, finding the sum of two points is rather easy - it may therefore come as a surprise that instead finding what number of times to add a point to itself in order to obtain a given other point is considered a very difficult problem to solve. This forms the basis of some of the most efficient and secure currently used cryptography methods.

Moduli spaces are important tools in solving classification problems and are subject to extensive study in abstract algebra and algebraic geometry. Their points represent mathematical objects, or oftentimes rather families of them that we consider to have identical properties. In such a space we may add coordinates, thus parameterizing the objects. Furthermore, additional structure such as topologies or operations may be embedded into the space, allowing for a more comprehensive study of its points.

Finding a moduli space of elliptic curves with a suitable topology would allow us to study in what way the properties of elliptic curves change when one curve is continuously

changed into another. We specifically consider elliptic curves over complex numbers, as they can be analyzed deeply and behave well under many of the operations we are interested in. In this paper, we find one potential moduli space of complex elliptic curves and show that it naturally allows for the definition of modular forms. We also show that modular forms do indeed exist through giving explicit construction. Finally, we demonstrate the power of modular forms by discussing a brief proof of Jacobi's Four Square Theorem.

The rest of this article is structured as follows. In section 2, we present some highly general concepts that are foundational to the topics we discuss. In section 3, we consider the underlying theory behind the objects we present, analyze and refer to throughout the article. In section 4 we investigate how to define a moduli space of elliptic curves, and we also visualize it for the sake of intuition. In section 5, we move on to studying the properties of modular forms and, more importantly, how they relate to each other. This theory is employed in section 6 to recover a quick proof of the Four-Square Theorem. Finally, we mention some final remarks and possible directions for future research in section 7.

2 Preliminaries

This section gives some general concepts that will be of great importance to the following sections. Note that this is only an outline of the theory that this article will use. Section 3 gives some more context, but the full details are still omitted for the sake of brevity. Serge Lang treats the topic comprehensively in [1].

A *group* is a set S along with a binary operation $+ : S^2 \to S$ satisfying three conditions:

1. *(Identity Element)* There exists an element $e \in S$ such that for any $a \in S$, $a + e = e + a = a$.

2. *(Inverse Element)* For any $a \in S$, there exists $a^{-1} \in S$ such that $a + a^{-1} = a^{-1} + a = e$.

3. *(Associativity)* For any $a, b, c \in S$ we have $(a + b) + c = a + (b + c)$.

If the binary operation also satisfies $a + b = b + a$ for all $a, b \in S$, the group is said to be *commutative* or *abelian*. A trivial example of an abelian group is the integers along with addition; an example of a non-abelian group is given by matrix multiplication discussed in section 3.5.

A *field* is a set S along with two binary operations, here denoted by $\cdot$ and $+$, satisfying conditions:

1. The $+$ and $\cdot$ operations, respectively, are commutative and associative.

2. The $+$ and $\cdot$ operations have identity elements denoted by 0 and 1, respectively.

3. *(Additive Inverse)* For any element $a \in S$, there exists an element $(-a) \in S$ such that $a + (-a) = (-a) + a = 0$.

4. *(Multiplicative Inverse)* For any element $a \in S\backslash\{0\}$, there exists an element $a^{-1} \in S$ such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

5. *(Distributivity of multiplication over addition)* For any $a, b, c \in S$ we have $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

A trivial example of a field is the set of real numbers with regular multiplication and addition; indeed, one may view groups and fields as generalizations of arithmetic on the real numbers, in order to allow similar expressions and calculations using the elements of other sets.

A $\mathbb{K}$-*vector space* is a set S along with a field $\mathbb{K}$ such that

1. Addition of elements in S is an abelian group.

2. For any $a, b \in \mathbb{K}$ and $\mathbf{v} \in S$, we have $a \cdot (b \cdot \mathbf{v}) = (a \cdot b) \cdot \mathbf{v}$.

3. If 1 is the multiplicative identity in $\mathbb{K}$, then $1 \cdot \mathbf{v} = \mathbf{v}\ \forall \mathbf{v} \in S$.

4. For any $a, b \in \mathbb{K}$ and $\mathbf{v} \in S$, we have $(a + b) \cdot \mathbf{v} = a \cdot \mathbf{v} + b \cdot \mathbf{v}$.

5. For any $a \in \mathbb{K}$ and $\mathbf{u}, \mathbf{v} \in S$, we have $a \cdot (\mathbf{u} + \mathbf{v}) = a \cdot \mathbf{u} + a \cdot \mathbf{v}$.

An *isomorphism* f between two sets A and B equipped with some structure is a structure-preserving invertible mapping. This essentially means that A and B are isomorphic if and only if their respective structures are identical. An *automorphism* is an isomorphism from some set to itself; it can be seen as a form of symmetry of the set (and indeed, symmetries of geometric objects are automorphisms).

A *lattice* $L = \langle z_1, z_2 \rangle$ is the set

$$\{(az_1 + bz_2) | a, b \in \mathbb{Z}\}$$

for two complex numbers z_1, z_2 which are linearly independent over $\mathbb{R}$. We say that z_1 and z_2 *generate* L.

A *homeomorphism* is a continuous bijective map with a continuous inverse. We view two homeomorphic objects as essentially the same, at least topologically speaking.

3 Theory

This section treats some of the more specific theory used throughout the article.

3.1 Moduli Space

A moduli space is a mathematical space whose points represent objects of a given kind. Such a space usually comes equipped with some topology, which gives a notion of when two objects of the space are similar. As classes of objects sometimes have natural isomorphisms (an example would be similarity for triangles), defining an equivalence relation on the space to identify certain objects makes sense. Put simply, we take objects to be distinct iff (if and only if) the isomorphisms cannot map them to each other. Topologically, the properties of the resulting quotient space are interesting since they are invariant under certain maps between mathematical objects.

To demonstrate, a somewhat trivial moduli space would be the moduli space of proper labeled triangles up to similarity. We may view such a triangle (up to similarity) as a triple (α, β, γ) of positive numbers representing the three angles of the triangle at vertices A, B, C. Thus, a moduli space of proper labeled triangles up to similarity is the set $\{(\alpha, \beta, \gamma) \in \mathbb{R}^3 | \alpha, \beta, \gamma > 0, \alpha + \beta + \gamma = \pi\}$. This space can naturally be equipped with the usual topology of $\mathbb{R}^3$, the Euclidean metric (or usual absolute value); we thus have a notion of two triangles being close together.

The above discussion lacks rigor - indeed, several definitions of a moduli space exist. The parametrization displayed in this paper will be referred to as a moduli space, although it technically fails the usual definitions due to the many automorphisms of complex elliptic curves. This will be of little importance to us, since endowing the set of our objects with

a tractable topological structure will more than suffice for our purposes of studying them.

3.2 Elliptic Curves

An *elliptic curve* E over a field $\mathbb{K}$ of characteristic not 2 or 3 is the set of solutions to an equation

$$y^2 = x^3 + Ax + B \tag{1}$$

where $A, B \in \mathbb{K}$. The equation (1) is called the *Weierstrass form* of the elliptic curve and is the form we will study in this article as it allows us to view E as a plane curve.

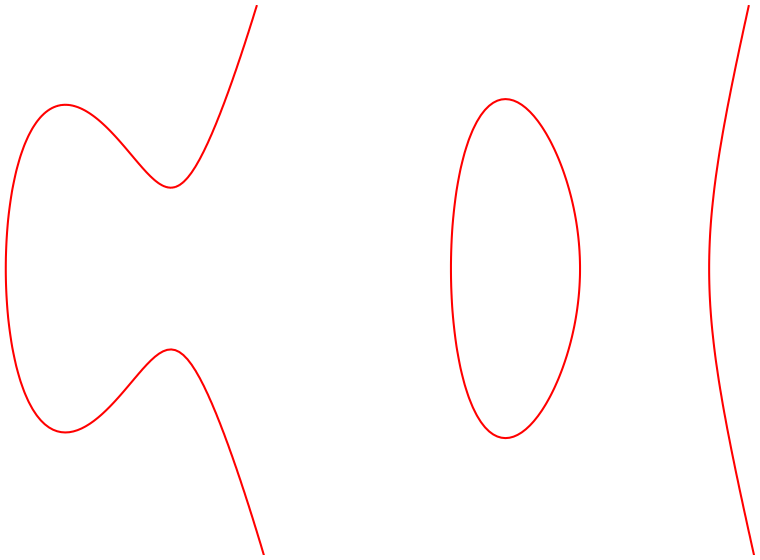

Figure 1: Two examples of elliptic curves, taken here to be defined and visualized over $\mathbb{R}$ for the sake of intuition (the closed loop is a part of the curve to the right).

The elliptic curve E also has a *point at infinity* O. It is a point of the *line at infinity*, which we in turn add so that we can say that there is precisely one line through any two points, and that two lines intersect in precisely one point. (Two curves *intersect* at a point if the point is a solution to the equations describing both curves.) This implies that an elliptic curve is defined over the *projective plane*, and lets us use a theorem due to Bézout which states that any line will intersect E precisely 3 times, counted with multiplicity (more generally, any two projective curves of degree m and n will intersect mn times, counting multiplicity). This will be important when we define addition on E.

A *singularity* of E is a point $P \in E$ such that any line intersecting E at P intersects E with multiplicity greater than 1. We require the discriminant

$$\Delta = 4A^3 + 27B^2 \tag{2}$$

to be nonzero, so that E has no singularities; the importance of this is shown in 3.3.

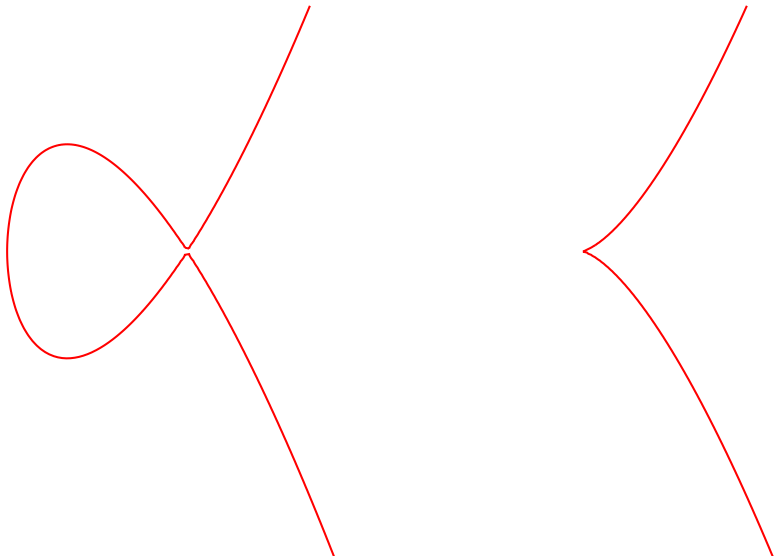

Figure 2: Two examples of elliptic curves with singularities. To the left, the singularity is the point where the elliptic curve self-intersects. To the right, the singularity is the sharp leftward protrusion.

3.3 Addition on Elliptic Curves

A group can be defined on E as follows: for distinct points P and Q, let R be the third intersection of the line PQ with E. If $P = Q$, let the line PQ be the tangent to E at P (the tangent exists as E is smooth; this follows from the condition $\Delta \neq 0$). We let $P + Q$ be the third point where the line OR intersects E. This is indeed a group: O is the identity, the third intersection between the line OP and E is $-P$, and it can be shown that associativity holds [2]. Furthermore, it is clear that $P + Q = Q + P$, so this group is also abelian.

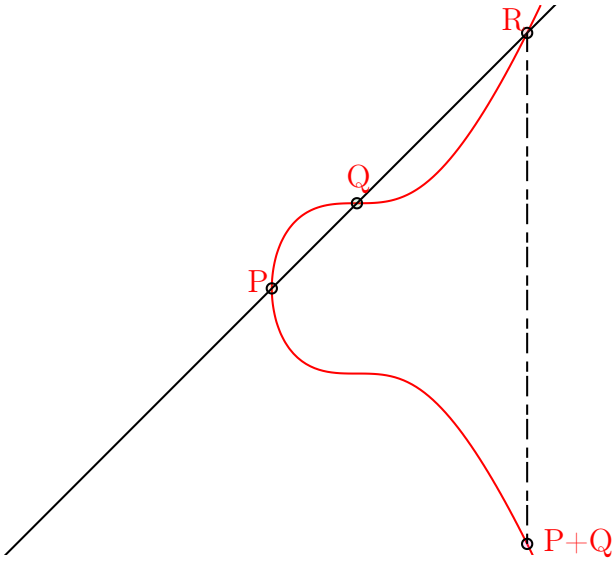

Figure 3: Addition of points P and Q on E. When defined over $\mathbb{R}$, an elliptic curve described by a Weierstrass equation will intersect the line at infinity at the point O where two vertical lines intersect. Therefore, the line through R and O will be the vertical line through R. Similarly, the point $P + Q$ is the reflection of R over the x-axis; since elliptic curves disregard the sign of y, they are always symmetric across the x-axis.

3.4 The j-Invariant

The *j-invariant* of the elliptic curve E is defined as

$$j = \frac{4A^3}{\Delta}. \tag{3}$$

There exists an isomorphism between two elliptic curves precisely when they have the same j-invariant [2], which will be important in constructing the moduli space of elliptic curves.

3.5 Modular Forms

We first define the group $SL_2(\mathbb{Z})$,which is the set of matrices

$$SL_2(\mathbb{Z}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : (a, b, c, d) \in \mathbb{Z}^4, ad - bc = 1 \right\} \tag{4}$$

equipped with regular matrix multiplication. Note that this is a noncommutative group. The group $GL_2(\mathbb{Z})$ which appears later is similarly defined; the only difference is that we

let $ad-bc=\pm 1$. Let $\mathcal{H}=\{a\in C:\operatorname{Im} a>0\}$ be the upper half of the complex plane, not including the reals. A *modular form* of *weight 2k* for $SL_2(\mathbb{Z})$ is a holomorphic function $f:\mathcal{H}\to\mathbb{C}$ such that $f(\gamma z)=(cz+d)^{2k}f(z)$ for all $\gamma\in SL_2(\mathbb{Z})$, where we define γz as $\frac{az+b}{cz+d}$. Such a function may be written as a function of $q=e^{2\pi iz}$ where $|q|<1$ [3]. We require $f(q)$ to have a value when $q=0$; this is equivalent to the existence of the limit of $f(z)$ when $q\to\infty$. We call this value the *value at infinity* of f.

A *cusp form* is a modular form that vanishes at infinity.

As we will show under section 5, the set M_{2k} of all modular forms for $SL_2(\mathbb{Z})$ of weight $2k$ can be interpreted as a $\mathbb{C}$-vector space of finite dimension.

3.6 Eisenstein Series

The *Eisenstein series of weight 2k* is defined as

$$G_{2k}(z)=\sum_{(m,n)\in\mathbb{Z}\setminus(0,0)}\frac{1}{(mz+n)^{2k}}. \tag{5}$$

It turns out that the Eisenstein series are not only modular forms, but basis elements for all modular forms for $SL_2(\mathbb{Z})$.

3.7 Weierstrass $\wp$ function

The *Weierstrass $\wp$ function* for a lattice L is defined as

$$\wp(z)=\frac{1}{z^2}+\sum_{\lambda\in L\setminus\{0\}}\left(\frac{1}{(z-\lambda)^2}-\frac{1}{\lambda^2}\right).$$

It is meromorphic; it is holomorphic away from poles of order 2 at every point of the lattice L. It is also doubly periodic; its two periods are the generators of L. [2]

4 Finding the Moduli Space

This section will present the main result of this article: a moduli space of complex elliptic curves up to group isomorphism.

4.1 The j-Invariant and a Trivial Moduli Space

An isomorphism class of elliptic curves (the set of all elliptic curves whose addition group is isomorphic to the addition group on a specified elliptic curve) is uniquely determined by its j-invariant. Conversely, the j-invariant determines a single isomorphism class of curves. Therefore, the j-invariant will be a strong tool in finding a moduli space. Indeed, it is easy to show that the j-invariant can assume any complex value; let $A = 1$ so that the j-invariant becomes $\frac{4}{4+B}$. This can assume any value but 0; to get 0, let $A = 0, B \neq 0$. Therefore, a trivial moduli space of complex elliptic curves up to isomorphism is $\mathbb{C}$.

This is indeed a parameterization of elliptic curves, and it does give us a concept of when two isomorphism classes of elliptic curves are close together. However, it does not permit much additional study. Our approach will therefore instead make use of the connection between lattices and elliptic curves to impose an equivalence relation on the upper half plane $\mathcal{H}$, identifying points z with their orbits under the $SL_2(\mathbb{Z})$ action. In particular, we will look at the fundamental domain. While the space thus created is homeomorphic to the trivial moduli space, it simultaneously serves as a parametrization of elliptic curves and the natural domain of modular forms for $SL_2(\mathbb{Z})$.

4.2 Lattices and Homotheties

The Weierstrass $\wp$ function gives an isomorphism of groups between lattices and elliptic curves; for any lattice L, it satisfies the differential equation

$$\wp'(z)^2 = 4\wp(z)^3 - 60G_4(L)\wp(z) - 140G_6(L). \tag{6}$$

It can be shown that taking $y = \wp', x = \wp$, this is a nonsingular elliptic curve. Conversely, every elliptic curve arises from some lattice. [2] The above equation can easily be put in Weierstrass form with $A = -15G_4(L), B = -35G_6(L)$. Due to this isomorphism, we may identify the set of nonsingular elliptic curves with the set of lattices over $\mathbb{C}$. Two lattices $\langle s, t\rangle$ and $\langle u, v\rangle$ define the same isomorphism class of elliptic curves iff they are homoth-

etic, that is, $\langle s,t\rangle = \lambda\langle u,v\rangle$ for some $\lambda \in \mathbb{C}\setminus\{0\}$. This follows simply from extending the definition of the Eisenstein series to the sum $\sum_{\gamma\in L\setminus\{0\}}\gamma^{-2k}$ for a lattice L. If $L_1 = \lambda L_2$, then $G_4(L_1) = \lambda^{-4}G_4(L_2)$ and $G_6(L_1) = \lambda^{-6}G_6(L_2)$. Letting this into the j-invariant, the powers of λ cancel. We omit the converse for brevity - the full details can be found in [2]. Thus, we may restrict ourselves to working with lattices $\langle 1,z\rangle$, and the homothety condition becomes $\langle 1,z\rangle = \lambda\langle 1,z'\rangle$ for some $\lambda\in\mathbb{C}\setminus\{0\}$. The following lemma elaborates on this.

Lemma. Lattices $\langle 1,z\rangle$ and $\langle 1,z'\rangle$ are homothetic iff $z' = \frac{az+b}{cz+d}$ with $ad-bc = \pm 1$.

Proof. If they are homothetic, then $\langle 1,z\rangle = \lambda\langle 1,z'\rangle$ for some $\lambda\in\mathbb{C}\setminus\{0\}$. Thus, $\lambda = cz+d$ and $\lambda z' = az+b$ for integers a,b,c,d, so

$$\begin{bmatrix}\lambda z'\\ \lambda\end{bmatrix} = \begin{bmatrix}a & b\\ c & d\end{bmatrix}\begin{bmatrix}z\\ 1\end{bmatrix}.$$

Note that this matrix is invertible; we could just as well write 1 and z as linear combinations of λ and λz since the lattices are the same. It is also of integer determinant, just like its inverse. Since $\det(AB) = \det(A)\det(B)$, both determinants are ± 1. This gives $ad-bc=\pm 1$. Taking $\frac{\lambda z'}{\lambda} = \frac{az+b}{cz+d} = z'$ yields one direction of the equivalence.

Conversely, if $z' = \frac{az+b}{cz+d}$ with $ad-bc=\pm 1$, then $\langle 1,z'\rangle = \langle 1,\frac{az+b}{cz+d}\rangle$. Multiplying by $cz+d$ gives a homothety to the lattice $\langle cz+d, az+b\rangle$. Taking $a(cz+d)-c(az+b) = ad-bc = 1$ and $-b(cz+d)+d(az+b) = (ad-bc)z = z$ gives that this is the lattice $\langle 1,z\rangle$, so that $\langle 1,z'\rangle$ is indeed homothetic to $\langle 1,z\rangle$. □

All lattices $\langle 1,z\rangle$ are parameterized by $\mathbb{C}$, but we have the equivalence relation $z\sim\frac{az+b}{cz+d}$; more formally, our space of lattices is $(\mathbb{C}\setminus\mathbb{R})/GL_2(\mathbb{Z})$ (two points of $\mathbb{C}\setminus\mathbb{R}$ are equivalent iff one can be taken to the other by a $GL_2(\mathbb{Z})$ matrix). It immediately follows that this is homeomorphic to $\mathcal{H}/SL_2(\mathbb{Z})$.

Lemma. $(\mathbb{C}\setminus\mathbb{R})/GL_2(\mathbb{Z})$ is homeomorphic to $\mathcal{H}/SL_2(\mathbb{Z})$.

Proof. The matrix taking z to $-z$ is in $GL_2(\mathbb{Z})$. We know that every homothety is given

by a $GL_2(\mathbb{Z})$ matrix. It remains to show that any homothety given by $GL_2(\mathbb{Z})$ action is also given by $SL_2(\mathbb{Z})$. We have $\langle 1, z\rangle = \langle cz + d, az + b\rangle$ for $ad - bc = -1$ and $\langle 1, z\rangle = \langle cz + d, a'z + b'\rangle$ for $a'd - b'c = 1$; dividing by $cz + d$ in both equations shows that the matrices

$$\begin{bmatrix} a' & b' \\ c & d \end{bmatrix} \text{ and } \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

give the the same homothety. □

The next lemma verifies our claim that this approach produces a space that is homeomorphic to the trivial moduli space.

Lemma. $\mathcal{H}/SL_2(\mathbb{Z})$ is homeomorphic to $\mathbb{C}$.

Proof. The function g sending points of $\mathcal{H}/SL_2(\mathbb{Z})$ to the j-invariants of their corresponding elliptic curves is bijective by the first lemma of this section. Using eq. (6) to obtain the elliptic curve corresponding to the lattice $\langle 1, z\rangle$, then taking it into Weierstrass form and writing the j-invariant gives that g is explicitly given by

$$g : z \mapsto \frac{20G_4^3(z)}{20G_4^3(z) + 49G_6^2(z)} \tag{7}$$

where G_4 and G_6 are the Eisenstein series of weights 4 and 6, respectively. Note that the denominator is nonzero as it is the discriminant of a nonsingular elliptic curve. Furthermore, Eisenstein series are modular forms (which will be discussed in section 5), thus holomorphic on $\mathbb{C}$. Therefore g is holomorphic. Now use Proposition 1.7.4 from [4]. It states that if $f : X \to Y$ is a surjective continuous map and we define an equivalence relation on X by saying $u \sim v$ iff $f(u) = f(v)$, then the induced map $f : X/\sim \ \to Y$ is a homeomorphism exactly when f sends saturated open sets $q^{-1}(U)$ to open sets. It also states that f sends saturated open sets to open sets if f is an open map.

Together with the Open Mapping Theorem from complex analysis which gives that g is indeed an open map, this yields that the induced map is a homeomorphism. □

4.3 Visualizing the Moduli Space

We now know that a moduli space of complex elliptic curves is $\mathcal{H}/SL_2(\mathbb{Z})$. We would however like a more concise description of the quotient. As a starting point for this, we find a generating set of $SL_2(\mathbb{Z})$.

4.3.1 Generators of $SL_2(\mathbb{Z})$

Lemma. A generating set of $SL_2(\mathbb{Z})$ is given by the two matrices

$$S = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \text{ and } T = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Proof. We first summarize the action of S and T.

$$S : \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \begin{bmatrix} a+c & b+d \\ c & d \end{bmatrix} \tag{8}$$

$$T : \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \begin{bmatrix} c & d \\ -a & -b \end{bmatrix} \tag{9}$$

The action of the respective inverses is clear. Our proof will proceed by starting with any matrix A in $SL_2(\mathbb{Z})$, and then showing that repeated application of S and T (or their inverses) reduces A to the identity matrix. We will work with the absolute values of a and c, using the following algorithm (and assuming all absolute values are greater than 1):

1. If $|a| > |c|$: Add/subtract c from a so that we have $a \to a'$ with $|a'| < |c|$.

2. If $|a| < |c|$: Apply T and then use step 1.

If $|a|$ or $|b|$ go below 2, then one of the absolute values has to be 1. As above, use this to achieve $(|a|, |c|) = (1, 0)$. We have now reduced the matrix A to a matrix A' of the form

$$A' : \begin{bmatrix} \pm 1 & b \\ 0 & \pm 1 \end{bmatrix} \tag{10}$$

where a has the same sign as d so the determinant equals 1. This reduces readily to the identity (or its negative) under repeated application of S, whence we may easily reduce to S and T. It is clear that $|a|+|c|$ is strictly decreasing under iterations of the algorithm, and therefore it must terminate. Inverting the sequence of matrices used, we have shown that any matrix of $SL_2(\mathbb{Z})$ can be generated from S and T. □

4.3.2 Fundamental Domain of $SL_2(\mathbb{Z})$

Having summarized the action of $SL_2(\mathbb{Z})$ on $\mathcal{H}$, we can now see precisely what $SL_2(\mathbb{Z})$ does to a point z in $\mathcal{H}$: its action is generated by $z \mapsto -\frac{1}{z}$ and $z \mapsto z+1$, along with the respective inverses. We shall now find a subset $\mathcal{F}$ of $\mathcal{H}$ such that any point of $\mathcal{H}$ can be sent into $\mathcal{F}$ under $SL_2(\mathbb{Z})$ action, and that no two points in $\mathcal{F}$ can be sent to each other; in other words, $\mathcal{F}$ is a *fundamental domain* of $SL_2(\mathbb{Z})$, and contains precisely one representative of every isomorphism class of complex elliptic curves (since it contains precisely one representative of each class of lattices up to homothety).

Lemma. A fundamental domain for $SL_2(\mathbb{Z})$ is given by

$\mathcal{F} = \{z \in \mathcal{H} | \, \mathrm{Re}(z) \in [-1/2, 1/2) \;\; \& \;\; |z| \geq 1 \;\; \text{with } |z| > 1 \text{ if } \mathrm{Re}(z) > 1\}$.

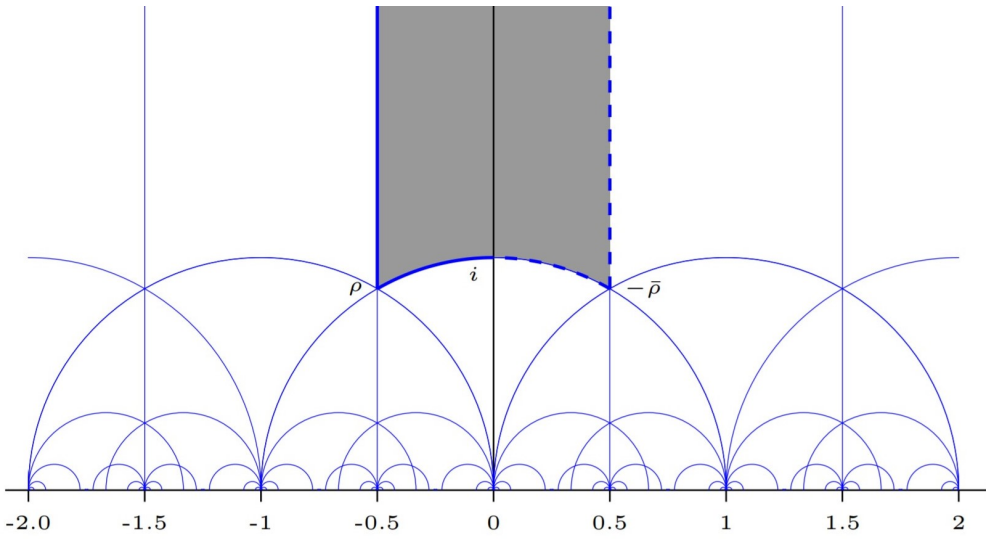

Figure 4: The set $\mathcal{F}$, which is a fundamental domain for $SL_2(\mathbb{Z})$.

Proof. We first prove that the orbit of any point intersects $\mathcal{F}$. Since we may translate

by 1 along the real axis, all orbits have points with real parts in the desired interval. We must now prove that the points with absolute value less than 1 can be made to have absolute value at least 1, while maintaining a real part within the desired interval. For each $0 < r < 1$, define $h_r(z)$ as the function whose domain is the semicircle of radius r in the upper half plane, and which lets $z \mapsto -\frac{1}{z}$ before adjusting the real part to lie in $[-1/2, 1/2)$. Note that if z is in the semicircle of radius r, letting z go to $h_r(z)$ multiplies the imaginary part of z by at least $\frac{1}{r}$, which is greater than 1. Therefore, suppose that there exists a point z_0 which stays in a semicircle of radius $r_0 < 1$ under applications of h. It then holds for every n that $\operatorname{Im} h^n_{r_0}(z) \geq 1/(r_0^n) \operatorname{Im} z$. Since $r_0 < 1$, the RHS can grow arbitrarily large, which contradicts the assumption that the absolute value of $h^n_{r_0}(z)$ is upper bounded by r_0. This shows that for any semicircle of radius $r < 1$, we can force all points of $\mathcal{H}$ to have a real part in the desired interval while having an absolute value greater than r; thus, application of S and T can force all points of $\mathcal{H}$ to have an absolute value of at least 1, and a real part in $[-1/2, 1/2)$. As to why the absolute value should be greater than 1 if the real part is positive, we can mirror the points of positive real part and absolute value of 1 across the y-axis by applying T. This shows that a point from every orbit is in $\mathcal{F}$. For the sake of brevity, the proof of the converse is taken from [3]. □

Thus, the fundamental domain $\mathcal{F}$ is a moduli space of complex elliptic curves. The reason why this particular visualization of the space is so interesting is that it is also the natural home of modular forms for $SL_2(\mathbb{Z})$. We shall now study these modular forms closer.

5 Modular Forms

We gave the definition and properties of modular forms in section 3.5. It is not trivial that such functions exist. We therefore give Eisenstein series as a demonstration; as stated in section 4.2, they are modular forms. We will now show this.

Lemma. Eisenstein series of weight $2k \geq 4$ are modular forms for $SL_2\mathbb{Z}$.

Proof. We shall first show that Eisenstein series respect the $SL_2\mathbb{Z}$ action; proving that the required properties hold under the application of the generating elements is sufficient.

Thus, what we need to check is $f(z+1) = z$ and $f\left(-\frac{1}{z}\right) = z^{2k} f(z)$. The former condition is obvious. For the latter, write

$$G_{2k}\left(\frac{-1}{z}\right) = \sum_{(m,n)\in\mathbb{Z}^2\setminus(0,0)} \frac{1}{(n-\frac{m}{z})^{2k}} = \sum_{(m,n)\in\mathbb{Z}^2\setminus(0,0)} \frac{z^{2k}}{(nz-m)^{2k}} = z^{2k} G_{2k}(z). \tag{11}$$

As to being holomorphic, it suffices to show that the summands of the Eisenstein series converge uniformly on compact subsets K of $\mathcal{H}$. We may thus assume $\operatorname{Im} z \geq m_0 > 0$ for some fixed m_0. Consider the *central parallelogram* P_1 about 0, being the set $\{1, -1, \pm z, 1 \pm z, -1 \pm z\}$. Our main idea is that the absolute values of the elements of P_1 are lower bounded by m_0 if $m_0 \leq 1$, and their contributions to the sum are thus upper bounded by $\frac{1}{m_0^{2k}}$. Further define $P_n = \{\{p \pm 1 \pm z\} : p \in P_{n-1}\}$ recursively, where we additionally require P_n to be disjoint from all earlier P_i.

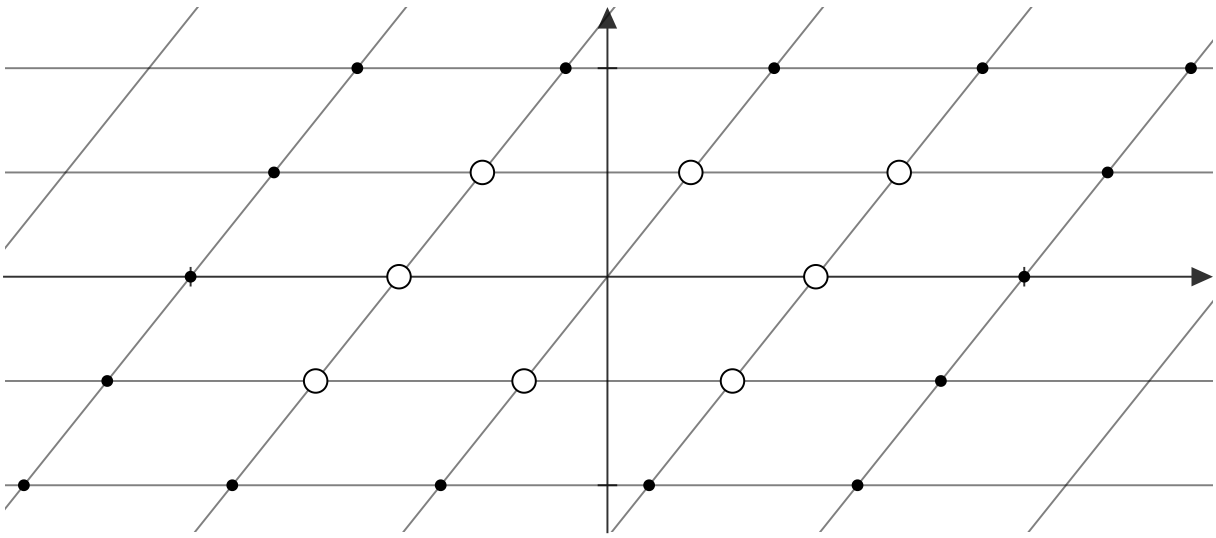

Figure 5: The central parallelogram P_1 for a lattice L, given by the eight encircled lattice points. The points marked in black form P_2.

The parallelogram P_n contains $8n$ points whose absolute values are lower bounded by nm_0; the contribution to the sum from P_n is therefore upper bounded by $\frac{8}{m_0^{2k} n^{2k-1}}$. The Eisenstein series can be rewritten as

$$G_{2k}(z) = \sum_{n\geq 0} \sum_{\gamma_n \in P_n} \frac{1}{\gamma_n^{2k}}. \tag{12}$$

What we want to show is uniform convergence. We therefore prove that

$$\sum_{n\geq M}\sum_{\gamma_n\in P_n}\frac{1}{|\gamma_n^{2k}|} \tag{13}$$

becomes arbitrarily small with large enough M independent of $z \in K$. By the preceding discussion, we have

$$\sum_{n\geq M}\sum_{\gamma_n\in P_n}\frac{1}{|\gamma_n^{2k}|} \leq \frac{8}{m_0^{2k}}\sum_{n\geq M}\frac{1}{n^{2k-1}} \tag{14}$$

where $\frac{8}{m_0^{2k}}$ is constant, and it is well known that the sum is arbitrarily small for large enough M since $2k-1 \geq 3$ (for example, one could overestimate the sum using a simple integral). This shows that $G_{2k}(z)$ converges uniformly absolutely on K and is therefore holomorphic there, since all the summands are. Finally, we find the value at infinity. Whenever m is not 0, the absolute value of the denominator of the summand goes to infinity and we can therefore disregard that summand. Therefore we only consider the summands where $m = 0$; this becomes

$$2\sum_{n\in\mathbb{N}\setminus(0,0)}\frac{1}{n^{2k}} = 2\zeta(2k). \tag{15}$$

We have thus shown that the Eisenstein series are modular forms. □

5.1 Δ and All Other Modular Forms

When viewed as a $\mathbb{C}$-vector space, the set M_k of modular forms of weight $2k$ has a finite basis. The full proof of this makes use of tedious contour integrals which we omit. We instead provide a sketch, where the full details can be found in [3].

Lemma. The $\mathbb{C}$-vector space of modular forms of weight $2k$ has a basis of the monomials $G_2^a G_3^b$ where $a, b \geq 0$ and $2a + 3b = k$.

Sketch of proof. We define the space M_k^0 of cusp forms of weight $2k$, which is the kernel of the linear map $f \to f(\infty)$. Thus, $dim(M_k) \leq dim(M_k^0) + 1$. Since G_{2k} is nonzero at infinity, we have $M_k = M_k^0 \bigoplus \mathbb{C}G_{2k}$. Multiplication by Δ gives an isomorphism between

M_{k-6} and M_k^0 since Δ is a cusp form of weight 12, and it is nonzero everywhere else. The fact that Δ is nonzero follows from the aforementioned contour integrals - it also follows from them that there are no modular forms of nonpositive weights. This completes the proof: the dimension of M_1, M_2, M_3, M_4 and M_5 is 1, and multiplication by Δ inductively gives the dimension of all other M_i. The dimension of M_k can now be computed for $k \geq 0$ as

$$\begin{cases} \dim\ M_k = [k/6] \text{ if } k \equiv 1 \mod 6 \\ \dim\ M_k = [k/6] + 1 \text{ if } k \not\equiv 1 \mod 6 \end{cases} \tag{16}$$

where $[x]$ denotes the integral part of x. We now know the dimension of M_k, and it is the same as the number of monomials as stated in the preceding lemma. It remains to show that the monomials are linearly independent over $\mathbb{C}$. Assume that $G_2^a G_3^b = \lambda G_2^{a_0} G_3^{b_0}$, where $\lambda \neq 0$ and $a_0 < a$. Then $G_2^{a-a_0} G_3^{b-b_0} = \lambda$ when $G_3 \neq 0$. This is impossible since G_2 and G_3 have different zeroes, so the monomials are linearly independent. Therefore, they constitute a basis for M_k. □

Remark. With $x^2 = G_2$ and $y^3 = G_3$, the above lemma shows that the set of all modular forms for $SL_2\mathbb{Z}$ is isomorphic to the set of homogeneous polynomials in x^2 and y^3.

6 Modular Forms & Four-Square Theorem

It is well-known that every non-negative integer can be written as a sum of four squares of integers. The exact number of quadruples (a, b, c, d) so that $a^2 + b^2 + c^2 + d^2 = n$ for any non-negative integer n is also known. It is a result originally due to Jacobi; a revised proof can be found in [5]. We will show that this result is easy to derive from the theory of modular forms.

In order to do this, we define the θ series as

$$\theta(z, 4) = \sum_{n \geq 0} K_4(n) q^n, K_4(n) = \left|\{(a, b, c, d) \in \mathbb{Z}^4 | a^2 + b^2 + c^2 + d^2 = n\}\right| \tag{17}$$

where $q = e^{2i\pi z}$. We then show that it is indeed a modular form for $\Gamma_0(4)$, a group defined as

$$\Gamma_0(4) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z}) | c \equiv 0 \mod 4 \right\}, \tag{18}$$

and can thus be expressed as a linear combination of the basis elements of $\Gamma_0(4)$. Finally, we take the q-series of the basis elements and solve a linear equation system to obtain the coefficients of the θ series, thus answering the question posed.

6.1 Modular Forms for $\Gamma_0(\mathbb{Z})$

As with the group $SL_2(\mathbb{Z})$, we would like to concisely summarize the group action by finding a set of generators for $\Gamma_0(4)$.

Lemma. The three matrices

$$S = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, U = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \text{ and } V = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

generate $\Gamma_0(4)$.

Proof. We again summarize the action of the matrices.

$$S : \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \begin{bmatrix} a+c & b+d \\ c & d \end{bmatrix}, U : \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \begin{bmatrix} a & b \\ 4a+c & 4b+d \end{bmatrix} \text{ and } V : \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \begin{bmatrix} -a & -b \\ -c & -d \end{bmatrix}$$

Our proof proceeds similarly to how we found the generators of $SL_2(\mathbb{Z})$. Given a matrix we construct another algorithm, once again given that all absolute values are at least 2.

1. If $|a| > |c|$: subtract/add to a some multiples of c until $\frac{1}{2}|c| \geq |a|$.

2. If $\frac{1}{2}|c| < |a| < |c|$: proceed as in step 1 so that $\frac{1}{2}|c| > |a|$.

3. If $\frac{1}{2}|c| > |a|$: Add/subtract $4a$ so that $|c|$ becomes less than $|a|$.

4. If any absolute value becomes 1 or less: c is 0, which gives $|a| = 1$. Applying V if necessary gives $(a, c) = (1, 0)$. Application of S then yields the identity matrix. □

We can now proceed to show that the K series respects this group action.

Lemma. The θ series respects the $\Gamma_0(4)$ group action.

Proof. Once again it suffices to show that the θ series respects the action of the generator elements. As we have defined θ in the variable q, θ(z+1) = θ(z) holds. In order to show that θ respects U, note that $K(n, a+b) = \sum_{i+j=n} K(i,a)K(j,b)$. Note also that

$$\theta(n,a)\theta(n,b) = \sum_{n\geq 0}\sum_{i+j=n} K(i,a)K(j,b)q^n = \sum_{n\geq 0} K(n,a+b)q^n = \theta(n,a+b). \tag{19}$$

We want to show $\theta\left(z,4\right)\left(4z+1\right)^2 = \theta\left(\frac{z}{4z+1},4\right)$. Taking the fourth root, this is equivalent to $\theta(z)\sqrt{4z+1} = \theta\left(\frac{z}{4z+1}\right)$ by the above. Using the identity $\theta\left(\frac{-1}{4z},1\right) = \theta(z,1)\sqrt{-2iz}$ [6] we get $\theta\left(\frac{z}{4z+1}\right) = \theta\left(\frac{1}{4+\frac{1}{z}}\right) = \theta\left(\frac{-1}{4\left(-1-\frac{1}{4z}\right)}\right) = \theta\left(-1-\frac{1}{4z}\right)\sqrt{-2i\left(-1-\frac{1}{4z}\right)} = \theta\left(\frac{-1}{4z}\right)\sqrt{2i+\frac{i}{2z}}$ which using the identity again is $\theta(z)\sqrt{2i+\frac{i}{2z}}\sqrt{-2iz} = \theta(z)\sqrt{4z+1}$. Taking the fourth powers of both sides yields that θ does indeed respect the $\Gamma_0(4)$ action. □

We are now in a position to prove that θ(n, 4) is a modular form.

Lemma. θ(n, 4) is a modular form for $\Gamma_0(4)$.

Proof. By definition, the value of θ at infinity is $K(0,4) = 1$. It remains to show that θ is holomorphic. We do this by proving that the q-series converges uniformly to θ given any positive lower bound on the imaginary part of the domain of θ.

First, note that all $K(n,4) \leq (2n+1)^4$. Let $y_0 > 0$ be our lower bound on the domain of θ. Rewrite the θ series as

$$\sum_{\substack{n\in\mathbb{Z}^4\\ |n|\geq 0}} e^{2\pi i|n|^2 z}. \tag{20}$$

We will show that the tail of the series

$$\sum_{\substack{n\in\mathbb{Z}^4\\ |n|\geq M_0}} \left|e^{2\pi i|n|^2 z}\right| \tag{21}$$

is arbitrarily small for large enough M independent of z. Note that

$$\left|e^{2\pi i|n|^2 z}\right| = \left|e^{2\pi i|n|^2 a}\right|\left|e^{-2\pi|n|^2 b}\right| \tag{22}$$

where a and b are the real and imaginary parts of z. The first of the two factors has a pure imaginary exponent, and its absolute value is therefore 1. The second factor has a negative real exponent; its absolute value is largest when the expression after the negative sign is as small as possible. The only part of the exponent that may vary is b; we therefore have

$$\left|e^{-2\pi\left|n^2 b\right|}\right| \leq \left|e^{-2\pi|n|^2 y_0}\right|. \tag{23}$$

Using this, we have

$$\sum_{\substack{n\in\mathbb{Z}^4 \\ |n|\geq M_0}} \left|e^{2\pi i|n|^2 z}\right| \leq \sum_{\substack{n\in\mathbb{Z}^4 \\ |n|\geq M_0}} e^{-2\pi|n|^2 y_0} \leq \sum_{m\geq M_0} (2m+1)^4 e^{-2\pi m y_0}. \tag{24}$$

We may assume that M_0 is a square. Then, by bunching terms together and rounding absolute values down to the nearest square, we have

$$\sum_{m\geq M_0} (2m+1)^4 e^{-2\pi m y_0} \leq \sum_{m\geq\sqrt{M_0}} (2m+1)(2(m+1)^2+1)^4 e^{-2\pi m^2 y_0} \tag{25}$$

which we can overestimate for large enough M_0 with

$$\sum_{m\geq\sqrt{M_0}} e^{m-2\pi m^2 y_0} \leq \sum_{m\geq\sqrt{M_0}} e^{-m} \tag{26}$$

which becomes arbitrarily small for large enough M_0, independent of the choice of z. The fact that θ is holomorphic above any given horizontal line in $\mathcal{H}$ gives that θ must be holomorphic at any point in $\mathcal{H}$, thus on all of $\mathcal{H}$. □

6.2 Answering the Question: Finding the Coefficients

It can be proven [6] that the $\mathbb{C}$-vector space $M_2(\Gamma_0(4))$ of modular forms of weight 2 for $\Gamma_0(4)$ has a basis

$$G_{2,2} = -\frac{\pi^2}{3}\left(1 + 24\sum_{n=1}^{\infty}\left(\sum_{\substack{0<d|n \\ \text{d odd}}} d\right)q^n\right)$$

$$G_{2,4} = -\pi^2\left(1 + 8\sum_{n=1}^{\infty}\left(\sum_{\substack{0<d|n \\ 4\nmid d}} d\right)q^n\right).$$

The series θ is a linear combination of these basis elements, i.e. $\theta = \lambda_1 G_{2,2} + \lambda_2 G_{2,4}$. Knowing that the first two coefficients of the q-expansion of θ are 1 and 8, we have the system of linear equations

$$\begin{cases} 1 = & -\frac{\pi^2}{3}\lambda_1 - \pi^2\lambda_2 \\ 8 = & -8\pi^2\lambda_1 - 8\pi^2\lambda_2 \end{cases}. \tag{27}$$

Dividing the second equation by 8 and subtracting from the first equation yields $0 = \frac{2\pi^2}{3}\lambda_1$ so that θ is a scalar multiple of $G_{2,4}$; matching the first coefficient of both series tells us that this scalar is $-\frac{1}{\pi^2}$. Thus, the expansion of the θ series is

$$\theta = 1 + 8\sum_{n=1}^{\infty}\left(\sum_{\substack{0<d|n \\ 4\nmid d}} d\right)q^n \tag{28}$$

and the number of ways in which the positive integer n can be written as a sum of four squares is therefore

$$8\sum_{\substack{0<d|n \\ 4\nmid d}} d, \tag{29}$$

which is indeed the result of Jacobi.

7 Concluding Remarks

In this article, we show that the set of complex elliptic curves up to group isomorphism can be parameterized by $\mathcal{H}/SL_2(\mathbb{Z})$, which is in turn homeomorphic to $\mathbb{C}$. We also demonstrate the existence of modular forms for $SL_2(\mathbb{Z})$, and we show that all modular forms for $SL_2(\mathbb{Z})$ are homogeneous polynomials in the two first Eisenstein series. Finally, we give a proof of the Four-Square Theorem through applying modular forms.

Several further questions arise from the results in this article. One could generalize the four-square theorem to other numbers of squares. As seen in section 6, the coefficients of the q-expansion of a modular form can carry much information - for example, one could try to come up with a general expression for the coefficients of the Eisenstein series. As we may multiply and add formal series, the coefficients of powers/products or sums of such series could also be found. Finally, one could study the properties of modular forms over our parameterization of isomorphism classes of complex elliptic curves, or generalize this parameterization to elliptic curves over any field.

References

[1] S. Lang, *Algebra*, pp. 7, 84, 139, 10. Springer, 2002.

[2] J. Silverman, *The Arithmetic of Elliptic Curves*, pp. 51, 45, 165, 169–170, 172–173. Springer, 2009.

[3] J.-P. Serre, *A Course in Arithmetic*, pp. 80, 78, 85–88. Springer, 1973.

[4] T. Lawson, *Topology: A Geometric Approach*, p. 39. Oxford University Press, 2003.

[5] M. D. Hirschhorn, "A simple proof of Jacobi's four-square theorem," *Journal of the Australian Mathematical Society*, vol. 32, pp. 61–67, Jan. 1982.

[6] J. H. Bruiner, G. van der Geer, G. Harder, and D. Zagier, *The 1-2-3 of Modular Forms*, pp. 25, 27. Springer, 2008.