Introduction to Modular Forms

Trevor Johnson

March 2026

1 Introduction

Modular forms are a class of functions that satisfy an infinite number of recurrences and symmetries given by a type of group called a congruence subgroup. While the connections may not be obvious on the surface, modular forms are related to many fields of math.

This paper serves as an introduction to modular forms. There is far too much known about modular forms to fit within a single paper, so to allow more information, justifications may not be provided for definitions, some proofs will not be written. However, when one is not already included, a reference will be provided containing a proof. The main source for this paper, [**Loz11**], contains justifications for the definitions given.

2 Complex Analysis

This section will cover the elements of complex analysis important to modular forms, but it assumes knowledge of the basics (e.g. what a complex number is, how functions work on them, e.t.c.) and of real calculus.

Definition 1. The *derivative* of a function f that is defined in a neighborhood of a point z_0 is equal to

$$\lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h}$$

at z_0.

Just as with real functions, the derivative may not exist everywhere. While this definition is extremely reminiscent of the definition of a derivative in real calculus, it is more restrictive.

Definition 2. A function is *holomorphic* (or *analytic*, which is equivalent but typically defined differently) on an open set $S \subseteq \mathbb{C}$ if it is complex differentiable at every point in S. It is holomorphic on a non-open set $S \subset \mathbb{C}$ if it is holomorphic on an open set containing S.

Definition 3. A function is *meromorphic* on S if it is holomorphic on S, excluding a set of isolated points.

3 Modular Curves

Definition 4. The *upper half-plane* $\mathbb{H} := \{a + bi : a, b \in \mathbb{R}, b > 0\}$ is the subset of the complex plane with positive imaginary coordinates. For any $z \in \mathbb{H}$ and $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2, \mathbb{Z})$, we define $Mz = \frac{az+b}{cz+d}$.

This action of the special linear group on $\mathbb{H}$ will allow us to create constructs, such as functions or groups, based on a combination of the two.

Definition 5.

$$\Gamma_0(N) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2,\mathbb{Z}) : c \equiv 0 \mod N \right\}$$
$$\Gamma_1(N) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2,\mathbb{Z}) : c \equiv 0, a \equiv d \equiv 1 \mod N \right\}$$
$$\Gamma(N) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2,\mathbb{Z}) : b \equiv c \equiv 0, a \equiv d \equiv 1 \mod N \right\}$$

A group G is a congruence subgroup if $\Gamma(N) \leq G \leq \mathrm{SL}(2,\mathbb{Z})$ for some $N \in \mathbb{Z}^+$.

Definition 6. Let Γ be a congruence subgroup. The points $\tau, \tau' \in \mathbb{H}$ are *equivalent relative to* Γ if there is a matrix $M \in \Gamma$ such that $\tau' = M\tau$. The set of equivalence classes is $Y(\Gamma) = \mathbb{H}/\Gamma$.

We will not prove this, as topology is outside the scope of this paper, but $Y(\Gamma)$ is not compact. To make $Y(\Gamma)$ compact, we need to add in some points. This requires us to extend $\mathbb{H}$ as Y is defined as a quotient.

Definition 7. *Homogeneous coordinates* are a set of coordinates $[x_0, x_1, \ldots, x_n]$. The difference between homogeneous coordinates and regular coordinates is that, for any $\lambda \neq 0$,

$$[x_0, x_1, \ldots, x_n] = [\lambda x_0, \lambda x_1, \ldots, \lambda x_n].$$

Definition 8. The *projective line* $\mathbb{P}^1(F)$ of a field F is the set of equivalence classes of homogeneous coordinates $[x_0, x_1]$ for $x_0, x_1 \in F$.

For all $x_1 \neq 0$, $[x_0, x_1] = [x_0 x_1^{-1}, 1]$, which maps to $x_0 x_1^{-1} \in F$. When $x_1 = 0$, we have the equivalence class of $[1, 0]$, which is typically referred to as the point at infinity, or simply ∞. Thus, $\mathbb{P}^1(F) = F \cup \{\infty\}$.

Definition 9. The extended upper half-pane is defined as $\mathbb{H}^* := \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$. For any $[s,t] \in \mathbb{P}^1(\mathbb{Q})$ and $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2,\mathbb{Z})$, we define

$$M[s,t] := [as + bt, cs + dt].$$

Note that for $s \in \mathbb{Q}$, $Ms = M[s,1] = [as + b, cs + d] \sim [\frac{as+b}{cs+d}, 1] = \frac{as+b}{cs+d}$, which matches our previous definition.

Definition 10. We can extend the equivalence that we defined in $\mathbb{H}$ to $\mathbb{H}^*$. Let Γ be a congruence subgroup. The points $\tau, \tau' \in \mathbb{H}^*$ are equivalent relative to Γ if there is a matrix $M \in \Gamma$ such that $\tau' = M\tau$. The set of equivalence classes $X(\Gamma) = \mathbb{H}^*/\Gamma$ is called a *modular curve.* The cusps of $X(\Gamma)$ are the equivalence classes with a representative in $\mathbb{P}^1(\mathbb{Q})$.

$X(\Gamma)$ is compact due to the added points, but we will not prove this.

Remark 1. For the specific cases of $\Gamma_0(N)$, $\Gamma_1(N)$, and $\Gamma(N)$, the modular curve is typically denoted as $X_0(N)$, $X_1(N)$, or $X(N)$, respectively.

4 Modular Form Basics

Definition 11. Let Γ be a convergence subgroup. A function $f : \mathbb{H} \to \mathbb{C}$ is weakly modular of weight k for Γ if

1. f is meromorphic on $\mathbb{H}$.
2. For any $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma$, $f(Mz) = (cz + d)^k f(z)$.

f is a modular function of weight k if it is weakly modular of weight k and is meromorphic at the cusps of the modular curve $X(\Gamma)$. If $\Gamma(N) \subseteq \Gamma$, we say the level of f is N.

Remark 2. The second condition is what creates the infinite symmetries of modular forms. For example, when $\Gamma = \mathrm{SL}(2,\mathbb{Z})$ (i.e. $N = 1$), $f(z+1) = f(z)$ and $f(-\frac{1}{z}) = z^k f(z)$.

Definition 12. $f : \mathbb{H} \to \mathbb{C}$ is a modular form of weight k for Γ if it is a modular function of weight k for Γ and is holomorphic on $\mathbb{H}$ and the cusps of the modular curve $X(\Gamma)$. A modular form is a cusp form if it vanishes at all of the cusps of $X(\Gamma)$.

Proposition 1. *Let $k, k' \leq 2$ be integers, and let Γ be a congruence subgroup.*

1. *Let f, g be modular forms of weight k for Γ. For all $\lambda, \mu \in \mathbb{C}$, $\lambda f(z) + \mu g(z)$ is also a modular form of weight k for Γ. (In other words, the set of modular forms of weight k for Γ is a vector space over $\mathbb{C}$.)*
2. *The set of cusp forms of weight k for Γ is a $\mathbb{C}$-linear subspace of the vector space of modular forms of weight k for Γ.*
3. *Let f, g be modular forms for Γ of weight k and k', respectively. $f(z)g(z)$ is a modular form of weight $k + k'$ for Γ.*

Proof. Since modular forms for Γ are holomorphic over $\mathbb{H}$ and $X(\Gamma)$ by definition, we will not need prove meromorphicity.

1. Let f be a modular form of weight k for Γ, let $\lambda \in \mathbb{C}$, and let $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma$. Scaling a holomorphic function by a constant factor will create another holomorphic function. This follows from the definition of complex differentiation, as the constant can simply be taken out of the limit. $\lambda f(Mz) = \lambda(cz + d)^k f(z) = (cz+d)^k \lambda f(z)$. Therefore, λf is a modular form of weight k for Γ.

   Now, let g also be a modular form of weight k for Γ. The fact that $(h + k)'(z) = h'(z) + k'(z)$ for any $h, k : \mathbb{H} \to \mathbb{C}$ complex differentiable at z can be easily derived from the definition of complex differentiation. This tells us that the sum of two holomorphic functions is also holomorphic. $(f + g)(Mz) = f(Mz) + g(Mz) = (cz+d)^k f(z) + (cz+d)^k g(z) = (cz+d)^k (f+g)(z)$, so $f + g$ is a modular form of weight k for Γ.

   Thus, the set of modular forms of weight k for Γ is a vector space over $\mathbb{C}$.
2. The sum of two functions that vanish at a point will also vanish at that point, and scaling a function that vanishes at a point by a constant factor will create a function that also vanishes at that point. Thus, the set of cusp forms of weight k for Γ is a $\mathbb{C}$-linear subspace of the vector space of modular forms of weight k for Γ.
3. Let f and g be modular forms of weight k and k', respectively, for Γ, and let $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma$. To figure out if fg is holomorphic, we will use the complex differentiation product rule $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$. The derivation for this is the same as the real product rule. Since f and g are both holomorphic, it follows that fg is holomorphic on $\mathbb{H}$ and the cusps of $X(\Gamma)$. $(fg)(Mz) = f(Mz)g(Mz) = (cz+d)^k f(z)(cz+d)^{k'} g(z) = (cz+d)^{k+k'}(fg)(z)$, so fg is a modular form of weight $k + k'$ for Γ. □

Definition 13. $M_k(\Gamma)$ refers to the $\mathbb{C}$-vector space of modular forms of weight k for Γ, and $S_k(\Gamma)$ refers to the subspace of cusp forms.

Proposition 2. *Let Γ, Γ' be congruence subgroups. If $\Gamma' \leq \Gamma$, then $M_k(\Gamma)$ is a $\mathbb{C}$-linear subspace of $M_k(\Gamma')$, and $S_k(\Gamma)$ is a $\mathbb{C}$-linear subset of $S_k(\Gamma')$.*

Proof. Let f be a modular form of weight k for Γ. If the recurrence $f(Mz) = (cz + d)^k f(z)$ is satisfied for all $M \in \Gamma$, then it is clearly satisfied for all $M \in \Gamma' \leq \Gamma$. From the way we created cusps, it follows that f is also holomorphic on the cusps of $X(\Gamma')$. Since we already know $M_k(\Gamma)$ is a vector space over $\mathbb{C}$, this tells us that $M_k(\Gamma)$ is a $\mathbb{C}$-linear subspace of $M_k(\Gamma')$. Adding in the additional condition of vanishing at the cusps makes it clear that $S_k(\Gamma)$ is a $\mathbb{C}$-linear subspace of $S_k(\Gamma')$. □

Definition 14. Let $k \geq 2$ be an integer, and let L be a lattice. The Eisenstein series of L of weight k is defined as

$$G_{2k}(L) = \sum_{0 \neq w \in L} \frac{1}{w^{2k}}$$

For $z \in \mathbb{H}$, we define $G_{2k}(z) = G_{2k}(\langle z, 1 \rangle)$.

Proposition 3. *$G_{2k}(z)$ is a modular form of weight $2k$ for $SL(2, \mathbb{Z})$. The value of G_{2k} at ∞ is $2\zeta(2k)$.*

For an outline of the proof, see Exercise 4.5.3 of [**Loz11**].

Definition 15. The *q-expansion* of a function $f(z)$ is

$$\sum_{n=-\infty}^{\infty} a_n q^n$$

for some sequence a_n.

Unless otherwise stated, we will be using $q = e^{2\pi i z}$ for all of our q-expansions. This is a special case called a Fourier expansion.

A function being holomorphic at ∞ is equivalent to the q-expansion having no terms with $n < 0$. To show that a weakly modular function of weight k is holomorphic on all of the cusps of $X(\Gamma)$, we can show that the q-expansion of $f[M]_k(z) = (cz + d)^{-k} f(Mz)$ has no terms with $n < 0$ for any $M \in \Gamma$.

This function notation will be useful later on, as $f[M_1 M_2]_k(z) = f[M_2]_k[M_1]_k(z)$. This will allow us to more easily test the modularity of functions.

Definition 16. A modular form is *normalized* if the first non-zero coefficient of its q-expansion is 1.

Proposition 4. *Let $k \geq 2$, let $\sigma_k(n) = \sum_{0<d|n} d^k$, and let B_n be the nth Bernoulli number. The q-expansion of the Eisenstein series is*

$$G_{2k}(z) = 2\zeta(2k) + \frac{2(2\pi i)^{2k}}{(2k-1)!} \sum_{n \geq 1} \sigma_{2k-1}(n) q^n$$

and the normalized Eisenstein series is

$$E_{2k}(z) = \frac{1}{2\zeta(2k)} G_{2k}(z) = 1 - \frac{4k}{B_{2k}} \sum_{n \geq 1} \sigma_{2k-1}(n) q^n$$

A proof of this can be found in Chapter 3, Proposition 6 of [**Kob93**].

Definition 17.

$$\Delta(z) = -16(4(-15G_4(z))^3 + 27(-35G_6(z))^2)$$

is a modular form of weight 12 for $\mathrm{SL}(2, \mathbb{Z})$ called the *modular discriminant.*

Theorem 1. *The dimension of $M_k(SL(2, \mathbb{Z}))$ is*

$$dim_{\mathbb{C}}(M_k(SL(2, \mathbb{Z}))) = \begin{cases} 0 & k < 0 \text{ or } k \text{ is odd} \\ \lfloor \frac{k}{12} \rfloor & k \geq 0 \text{ and } k \equiv 2 \mod 12 \\ \lfloor \frac{k}{12} \rfloor + 1 & \text{otherwise} \end{cases}$$

In addition, for any $k \in \mathbb{Z}$, there is an isomorphism ψ between $M_{2k-12}(SL(2, \mathbb{Z}))$ and $S_{2k}(SL(2, \mathbb{Z}))$ given by $\psi(f(z)) = \Delta(z) f(z)$.

The proof of this theorem can be found in Chapter 7, Theorem 4 of [**Ser73**]. From this, it is clear that modular forms share a deep relation to the Riemann zeta function. For better or for worse, it is only related to $\zeta(2k)$, which is already well understood. However, this paper only looks at modular forms of integer weight, and there may be connections with better application for non-integer weight.

5 Math Applications

The proof of Fermat's Last Theorem involves modular forms. The statement of Fermat's Last Theorem is that, for any integer $n > 2$, there are no nonzero $a, b, c \in \mathbb{Z}$ such that $a^n + b^n = c^n$. Fermat claimed to have a proof too large to fit in the margin, but it took 358 years before it was proved by a mathematician named Andrew Wiles. A previous mathematician named Gerhard Frey conjectured that if Fermat's Last Theorem was false, the Taniyama-Shimura-Weil Conjecture, a theorem relating elliptic curves to modular forms, would be false. Wiles was able to prove the Taniyama-Shimura-Weil conjecture, proving Fermat's Last Theorem. The proof is around 100 pages long and can be read at [**Wil95**]

Another theorem proved with modular forms is the Jacobi Four Square Theorem, the proof of which is much shorter.

Theorem 2 (Jacobi Four Square Theorem)**.** *The total number of ways to write an integer n as the sum of 4 squares is* $r_4(n) = \sum_{d|n, d \notin 4\mathbb{Z}} 8d$.

Definition 18. $r_k(n)$ is the number of ways to write n as the sum of k squares. More precisely, $r_k(n) = \#\{(a_1, a_2, \ldots, a_k) \in \mathbb{Z}^k | \sum a_i^2 = n\}$.

Definition 19. $\theta : \mathbb{H} \to \mathbb{C}$ is defined as

$$\theta(z) = \sum_{m \in \mathbb{Z}} e^{2\pi i z m^2}.$$

The absolute value of $\theta(z)$ is bounded by $2 \sum_{n \geq 0} e^{-2\pi Im(z) n} = \frac{2}{1 - e^{-2\pi Im(z)}}$. This bound converges since $Im(z) > 0$. Since the terms of θ go to 0 as $|m| \to \infty$, θ also converges. Now, we will connect θ to r.

$$\theta(z)^k = \sum_{n \in \mathbb{Z}} \left(\sum_{(a_1, \ldots, a_k), \sum a_i^2 = n} 1 \right) q^n = \sum_{n \in \mathbb{Z}} r_k(n) q^n$$

This tells us that we should care about $\theta(z)^4$.

Theorem 3. $\theta(z)^4 \in M_2(\Gamma_0(4))$.

Lemma 1. $\Gamma_0(4)$ *is generated by* $\pm \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ *and* $\pm \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}$.

Proof. Let $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $c = 0$, then $a = d = \pm 1$, so M is generated by $\pm \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Suppose $c \neq 0$. We have, where $*$ represents a complicated value we do not care about:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^n = \begin{bmatrix} a & * \\ c & nc + d \end{bmatrix}$$
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}^n = \begin{bmatrix} * & b \\ c + 4md & d \end{bmatrix}$$

The first allows us to choose n such that $|nc + d| < |c|/2$. The second allows us to choose n such that $|c + 4md| < 2d$. Thus, we can repeatedly reduce $|c|$ and $|d|$. Since d could never be 0, eventually c will equal 0. ☐

Proof of Theorem 3.

$$\theta(z) = \sum_{m \in \mathbb{Z}} e^{2\pi i z m^2} = \sum_{m \in \mathbb{Z}} e^{2\pi i z m^2 + 2\pi i m^2} = \theta(z + 1)$$

This tells us that $\theta(z)^4 = \theta(z+1)^4$. Let $f(x) = e^{-\pi t x^2}$.

$$\begin{aligned}\hat{f}(n) &= \int_{-\infty}^{\infty} e^{-\pi t x^2 - 2\pi x n}\,dx = e^{\frac{\pi n^2}{t}} \int_{-\infty}^{\infty} e^{-\pi t (x - \frac{ni}{t})^2}\,dx \\ &= e^{\frac{\pi n^2}{t}} \int_{-\infty}^{\infty} e^{-\pi t x^2}\,dx = \frac{1}{\sqrt{t}} e^{\frac{\pi n^2}{t}}\end{aligned}$$

Let $z = \frac{-t}{2i}$. By the Poisson Summation Formula, $\theta(-\frac{1}{4z}) = \sum_{m\in\mathbb{Z}} e^{\frac{\pi n^2}{t}} = \sqrt{t}\sum_{m\in\mathbb{Z}} e^{\pi t n^2} = \sqrt{-2iz}\theta(z)$. Thus,

$$\begin{aligned}\theta\left(\frac{z}{4z+1}\right)^4 &= \theta\left(-\frac{1}{4(-\frac{1}{4z}-1)}\right)^4 = \left(2i\left(\frac{1}{4z}+1\right)\right)^2 \theta\left(-\frac{1}{4z}-1\right)^4 \\ &= \left(2i\left(\frac{1}{4z}+1\right)\right)^2 \theta\left(-\frac{1}{4z}\right)^4 \\ &= \left(2i\left(\frac{1}{4z}+1\right)\right)^2 (-2iz)^2 \theta\left(z\right)^4 \\ &= (4z+1)^2 \theta\left(z\right)^4\end{aligned}$$

Therefore, $\theta(z)^4$ is a modular form of weight 2 for $\Gamma_0(4)$. $\square$

Now, we will figure out the dimension of $M_2(\Gamma_0(4))$. This will allow us to create a basis and find the value of $\theta(z)^4$ in terms of that basis. To do this, we will establish a new and equivalent definition for modular forms of weight $2k$.

Definition 20. Let $k \in \mathbb{Z}$. A holomorphic function $f : \mathbb{H} \to \mathbb{C}$ is a *modular form* of weight $2k$ for Γ if:

1. $f(z)(dz)^k = f(Mz)(d(Mz))^k$ for all $M \in \Gamma$.
2. f is holomorphic at the cusps of the modular curve $X(\Gamma)$.

This clearly satisfies every condition except the recurrence condition, which we need to prove. $d(Mz) = d(\frac{az+b}{cz+d}) = \frac{ad-bc}{(cz+d)^2}\,dz = \frac{1}{(cz+d)^2}\,dz$, so the recurrence is satisfied.

Theorem 4. $\dim M_2(\Gamma_0(4)) = 2$.

Proof. This proof requires some topology, which is outside the scope of this paper, but a summary will still be provided. The full proof can be found in Chapter 6, Section 1 of [**Sri19**]. First, we discover the infinite region bounded below by semicircles of radius $\frac{1}{4}$ at $\pm\frac{1}{4}$ and bounded on the sides by $|Re(z)| \leq \frac{1}{2}$ to be a fundamental domain of $\Gamma_0(4)$. Then, we show this to be equivalent to a sphere with punctures at 0, 1, and ∞. Next, we show that the space of 1-forms over this sphere is generated by $\frac{dz}{z}$ and $\frac{dz}{z-1}$. By our redefinition of modular forms of even weight, this space is equivalent to $M_2(\Gamma_0(4))$, so $\dim M_2(\Gamma_0(4)) = 2$. $\square$

The basis we are going to construct involves G_2. In our definition of the Eisenstein series, we started at G_4. This is not because G_2 does not exist, but because it is not a modular form. Let $\mathbb{Z}$ equal $\mathbb{Z}\backslash 0$ if $c = 0$ and $\mathbb{Z}$ otherwise. $G_2(z) = \sum_{c\in\mathbb{Z}} \sum_{d\in\mathbb{Z}_c} \frac{1}{(cz+d)^2}$. It is important to note that this does not converge absolutely, so we must treat it with care. We will rewrite G_2 with the help of these two identities:

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

$$\pi \cot \pi z = -\pi i - 2\pi i \sum_{n=0}^{\infty} q^n$$

Taking the derivative of these two identities gives a third identity, namely $\sum_{n=-\infty}^{\infty} \frac{1}{(z+n)^2} = -4\pi^2 \sum_{n=0}^{\infty} nq^n$. Using this, we can re-express G_2:

$$\begin{aligned} G_2(z) &= \sum_{c\in\mathbb{Z}} \sum_{d\in\mathbb{Z}_c} \frac{1}{(cz+d)^2} \\ &= \sum_{d\neq 0} \frac{1}{d^2} + \sum_{c>0} \sum_{d=-\infty}^{\infty} \frac{1}{(cz+d)^2} + \sum_{c<0} \sum_{d=-\infty}^{\infty} \frac{1}{(cz+d)^2} \\ &= 2\zeta(2) + 2\sum_{c>0} \sum_{d=-\infty}^{\infty} \frac{1}{(cz+d)^2} \\ &= 2\zeta(2) + 2\sum_{c>0}(-4\pi^2) \sum_{d=0}^{\infty} dq^{cd} \\ &= \frac{\pi^2}{3} - 8\pi^2 \sum_{n=1}^{\infty} \sigma(n)q^n \end{aligned}$$

where $\sigma(n) = \sum_{d|n} n$.

Theorem 5. *$G_2[M]_2(z) = G_2(z) - \frac{2\pi ic}{cz+d}$ for all $M \in SL(2,\mathbb{Z})$.*

Proof. We will first show that this is true for $M = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. In other words, we want to prove that $G_2(z) = G_2(z+1)$.

$$G_2(z+1) = 2\zeta(2) - 8\pi^2 \sum_{c>0} \sum_{n=0}^{\infty} de^{2\pi icd(z+1)} = 2\zeta(2) - 8\pi^2 \sum_{c>0} \sum_{n=0}^{\infty} de^{2\pi icdz} = G_2(z)$$

Now, we will show this to be true for the other generating matrix of SL$(2,\mathbb{Z})$, $M = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. In other words, we want to show that $z^{-2}G_2(-\frac{1}{z}) = G_2(z) - \frac{2\pi i}{z}$.

$$\begin{aligned} z^{-2}G_2\left(-\frac{1}{z}\right) &= z^{-2} \sum_{c\in\mathbb{Z}} \sum_{d\in\mathbb{Z}_c} \frac{1}{(-c/z+d)^2} \\ &= \sum_{c\in\mathbb{Z}} \sum_{d\in\mathbb{Z}_c} \frac{1}{(-c+dz)^2} \\ &= \sum_{d\in\mathbb{Z}} \sum_{c\in\mathbb{Z}_d} \frac{1}{(cz-d)^2} \\ &= \sum_{d\in\mathbb{Z}} \sum_{c\in\mathbb{Z}_d} \frac{1}{(cz+d)^2} \\ &= \sum_{d\neq 0} \frac{1}{d^2} + \sum_{d\in\mathbb{Z}} \sum_{c\neq 0} \frac{1}{(cz+d)^2} \end{aligned}$$

While this appears to just be $G_2(z)$, we must keep in mind that we cannot simply swap the sums as this does not converge absolutely. We will use $\sum_{d\in\mathbb{Z}} \frac{1}{(cz+d)(cz+d+1)}$, which is a telescoping series and equals 0, to express G_2 in a manner easier to apply here.

$$\begin{aligned} G_2(z) &= \frac{\pi^2}{3} + \sum_{c\neq 0} \sum_{d\in\mathbb{Z}} \frac{1}{(cz+d)^2} - \sum_{c\neq 0} \sum_{d\in\mathbb{Z}} \frac{1}{(cz+d)(cz+d+1)} \\ &= \frac{\pi^2}{3} + \sum_{c\neq 0} \sum_{d\in\mathbb{Z}} \frac{1}{(cz+d)^2(cz+d+1)} \end{aligned}$$

This new sum is absolutely convergent as the term being summed is on the order of $\frac{1}{(cz+d)^3}$, so we can swap the sums. Thus, we get that

$$G_2(z) = \frac{\pi^2}{3} + \sum_{d\in\mathbb{Z}}\sum_{c\neq 0}\frac{1}{(cz+d)^2} - \sum_{d\in\mathbb{Z}}\sum_{c\neq 0}\frac{1}{(cz+d)(cz+d+1)} = z^{-2}G_2\left(-\frac{1}{z}\right) - \sum_{d\in\mathbb{Z}}\sum_{c\neq 0}\frac{1}{(cz+d)(cz+d+1)}.$$

Now, we will show that $\lim\limits_{N\to\infty}\sum\limits_{d=-N}^{N-1}\sum\limits_{c\neq 0}\frac{1}{(cz+d)(cz+d+1)} = -\frac{2\pi i}{z}$. When N is finite, this converges absolutely, so we have

$$\begin{aligned}\sum_{d=-N}^{N-1}\sum_{c\neq 0}\frac{1}{(cz+d)(cz+d+1)} &= \sum_{c\neq 0}\sum_{d=-N}^{N-1}\frac{1}{(cz+d)(cz+d+1)}\\ &= \sum_{c\neq 0}\frac{1}{cz-N} - \sum_{c\neq 0}\frac{1}{cz+N}\\ &= -\frac{1}{z}\sum_{c\neq 0}\left(\frac{1}{N/z-c} + \frac{1/z}{N/z+c}\right)\\ &= \frac{2z}{N} - \frac{1}{z}2\pi\cot(\pi N/z)\end{aligned}$$

which tells us that

$$\begin{aligned}\lim_{N\to\infty}\sum_{d=-N}^{N-1}\sum_{c\neq 0}\frac{1}{(cz+d)(cz+d+1)} &= \lim_{N\to\infty}\left(\frac{2z}{N} - \frac{2\pi\cot(\pi N/z)}{z}\right)\\ &= -\lim_{N\to\infty}\frac{2\pi\cot(\pi N/z)}{z}\\ &= -\frac{2\pi}{z}\lim_{N\to\infty} i\frac{e^{2\pi i N/z}+1}{e^{2\pi i N/z}-1}\\ &= -\frac{2\pi i}{z}.\end{aligned}$$

We will now show that this is true for the inverses of these matrices. For the first generating matrix, the inverse is just a translating matrix in the opposite direction, and $G_2(z) = G_2((z-1)+1) = G_2(z-1)$. The second generating matrix is its own inverse. Now, we just need to show that our formula satisfies the multiplicity. Let $M_1 = \begin{bmatrix} a & b \\ c & d\end{bmatrix}, M_2 = \begin{bmatrix} e & f \\ g & h\end{bmatrix} \in \mathrm{SL}(2,\mathbb{Z})$.

$$\begin{aligned}f[M_1]_2[M_2]_2(z) &= \left(f - \frac{2\pi i c}{cz+d}\right)[M_2]_2(z)\\ &= f(z) - \frac{2\pi i g}{gz+h} - (gz+h)^{-2}\frac{2\pi i c}{c\frac{ez+f}{gz+h}+d}\\ &= f(z) - \frac{2\pi i c - 2\pi i g(c(ez+f)+d(gz+h))}{(gz+h)(c(ez+f)+d(gz+h))}\end{aligned}$$

We want to prove that $\frac{2\pi i c - 2\pi i g(c(ez+f)+d(gz+h))}{(gz+h)(c(ez+f)+d(gz+h))} = \frac{2\pi i(ce+dg)}{(ce+dg)z+(cf+dh)}$, so we will show that $2\pi i c + 2\pi i g(c(ez+f)+d(gz+h)) = 2\pi i(ce+dg)(gz+h)$. Expanding and canceling gives $2\pi i(c+gcf) = 2\pi i ceh$, which is true because $\det M_2 = 1$. $\square$

We want to modify G_2 to become a modular form, but we will first establish a theorem that will help us with this.

Theorem 6. *Let $f : \mathbb{H} \to \mathbb{C}$ be weakly modular for Γ with level N. If there are $C, r > 0$ such that the q-expansion of f is $f(x) = \sum_{n\geq 0} a_n e^{2\pi i n z/N}$ where $0 \leq a_n \leq Cn^r$, then*

$$|f(z)| \leq C_0 + C\left(\int_0^\infty t^r e^{-2\pi t Im(z)/N}\,dt\right) + \frac{C_1}{Im(z)^r}$$

for some C_1, C_2, and f is a modular form for Γ.

Proof. It is clear that $|f(z)| \leq |a_0| + \sum_{n>0} Cn^r e^{-2\pi n Im(z)/N}$. Let $g(t) = t^r e^{-2\pi t Im(z)/N}$. When $t \in (0, \frac{rN}{2\pi Im(z)}$, $g'(t) > 0$, and when $t > \frac{rN}{2\pi Im(z)}$, $g'(t) < 0$. This tells us that

$$\sum_{n=1}^{k-1} n^r e^{-2\pi n Im(z)} < \int_0^k t^r e^{-2\pi t y/N}\,dt \text{ and}$$
$$\sum_{n=k+2}^{\infty} n^r e^{-2\pi n Im(z)} < \int_k^\infty t^r e^{-2\pi t y/N}\,dt$$

where $k = \lfloor \frac{rN}{2\pi Im(z)} \rfloor$. Therefore,

$$\begin{aligned}|f(z)| &\leq |a_0| + C\left(k^r e^{-2\pi k Im(z)/N} + (k+1)^r e^{-2\pi(k+1)Im(z)/N} + \sum_{n=1}^{k-1} n^r e^{-2\pi n Im(z)} + \sum_{n=k+2}^{\infty} n^r e^{-2\pi n Im(z)}\right)\\ &\leq C_0 + \frac{C_1}{Im(z)^r} + C\left(\int_0^\infty t^r e^{-2\pi t Im(z)/N}\right).\end{aligned}$$

We have $C\left(\int_0^\infty t^r e^{-2\pi t Im(z)/N}\,dt\right) = \frac{C}{y^{r+1}}\left(\int_0^\infty t^r e^{-2\pi t/N}\right) = \frac{C_2}{y^{r+1}}$ for some C_2. As $Im(z) \to \infty$, we have:

$$\begin{aligned}|f[M]_k(z)| = |(cz+d)^{-k} f(Mz)| &\leq \left|(cz+d)^{-k}(C_0 + \frac{C_1}{(Im(Mz))^r} + \frac{C_2}{(Im(Mz))^{r+1}}\right|\\ &= \left|(cz+d)^{-k}(C_0 + \frac{C_1(cz+d)^{2r}}{Im(z)^r} + \frac{C_2(cz+d)^{2r+2}}{Im(z)^{r+1}}\right| = O(Im(z)^{r+1-k})\end{aligned}$$

Therefore, $\lim_{z\to\infty} |f[M]_k e^{2\pi i z/N}| = \lim_{Im(z)\to\infty} O(y^{r+1-k}) e^{-2\pi Im(z)/N} = 0$, so f is a modular form. ☐

Definition 21. We define $G_{2,N}(z) = G_2(z) - NG_2(Nz)$.

We will prove that $G_{2,N}$ is weakly modular of weight 2 for $\Gamma_0(N)$. Let $M = \begin{bmatrix} a & b \\ Nc & d \end{bmatrix}$ and $M' = \begin{bmatrix} a & Nb \\ c & d \end{bmatrix}$. $N(Mz) = M'(Nz)$. With this, we can calculate that

$$\begin{aligned}G_{2,N}(Mz) = G_2(Mz) - NG_2(N(Mz)) &= (Ncz+d)^2 G_2(z) - 2\pi Nc(Ncz+d) - NG_2(N(Mz))\\ &= (Ncz+d)^2 G_2(z) - 2\pi Nc(Ncz+d) - NG_2(M'(Nz))\\ &= (Ncz+d)^2 G_2(z) - 2\pi i Nc(Ncz+d) - N((Ncz+d)^2 G_2(Nz) - 2\pi i c(Ncz+d))\\ &= (Ncz+d)^2 (G_2(z) - NG_2(Nz)).\end{aligned}$$

Since $G_{2,N}$ is clearly meromorphic, it is weakly modular. Note that $\Gamma_0(4) \subseteq \Gamma_0(2)$, so $G_{2,2}$ is also weakly modular over $\Gamma_0(4)$. Now, we will apply theorem 6. The q-expansion of $G_{2,2}$ is

$$\begin{aligned}G_{2,2}(z) &= G_2(z) - 2G_2(2z)\\ &= 2\zeta(2) - 8\pi^2 \sum_{n=1}^{\infty} \sigma(n) q^n - 2\left(2\zeta(2) - 8\pi^2 \sum_{n=1}^{\infty} \sigma(n) q^{2n}\right)\\ &= -\frac{\pi^2}{3} - 8\pi^2 \sum_{n=1}^{\infty}\left(\sum_{d|n, d\notin 2\mathbb{Z}} d\right) q^n.\end{aligned}$$

Similarly, the q-expansion of $G_{2,4}$ is

$$\begin{aligned} G_{2,4}(z) &= G_2(z) - 4G_4(2z) \\ &= 2\zeta(2) - 8\pi^2 \sum_{n=1}^{\infty} \sigma(n) q^n - 4\left(2\zeta(2) - 8\pi^2 \sum_{n=1}^{\infty} \sigma(n) q^{4n}\right) \\ &= -\pi^2 - 8\pi^2 \sum_{n=1}^{\infty} \left(\sum_{d|n, d \notin 4\mathbb{Z}} d \right) q^n. \end{aligned}$$

Thus, $G_{2,2}, G_{2,4} \in M_2(\Gamma_0(4))$. Since $G_{2,4}$ is clearly not a constant multiple of $G_{2,2}$, these two functions form a basis of $M_2(\Gamma_0(4))$. Now, we will look at the first two terms of θ^4, $G_{2,2}$, and $G_{2,4}$:

$$\begin{aligned} \theta(z)^4 &= 1 + 8q + \dots \\ G_{2,2}(z) &= -\frac{\pi^2}{3}(1 + 24q + \dots) \\ G_{2,4}(z) &= -\pi^2(1 + 8q + \dots) \end{aligned}$$

Therefore, $\theta(z)^4 = -\frac{1}{\pi^2} G_{2,4}(z)$, so $r_4(n) = 8 \sum\limits_{d|n, d \notin 4\mathbb{Z}} d.$ ☐

6 Other Applications

Modular forms appear in String Theory, specifically relating to black holes. Jacob Bekenstein and Stephen Hawking discovered that black holes are thermodynamic objects, and therefore have entropy. Through thought experiments about scattering waves, Hawking was able to create this formula for the entropy of a black hole:

$$S_{\text{BH}}^{\text{class}} = k_B \frac{1}{4} \frac{A_H}{\ell_{\text{PL}}^2}$$

where $\ell_{\text{PL}}^2 = \frac{G\hbar}{c^3}$ and A_H is the area of the event horizon. Using Boltzmann's equation which relates the number of microstates to the entropy, this expression can also be found:

$$k_B \log d_{\text{micro}} = S_{\text{BH}}^{\text{class}}$$

This equation is approximate, but it is valid in the thermodynamic limit of black hole size. The number of microstates and the microscopic degeneracy of states have been discovered to be closely related to modular forms, further connecting anything related to black hole entropy or the area of the event horizon to modular forms. String Theory and Black Hole theory are outside the scope of this paper, so the connections will not be explained any more here. A more in-depth explanation can be found at [**Mur23**].

7 Conclusion

The information presented in this paper is merely an introduction, and modular forms run much deeper than shown here. They are connected to many different mathematical fields, and they may even be connected to the most fundamental building blocks of reality. It may not seem so initially, but modular forms are incredibly important.

References

[APP11] Ravi Agarwal, Kanishka Perera, Sandra Pinelas, *An Introduction to Complex Analysis*, Springer, New York, 2011.

[Hit03] Nigel Hitchin, *PROJECTIVE GEOMETRY: b3 course 2003*, 2003.

[Kob93] Neal I. Koblitz, *Introduction to Elliptic Curves and Modular Forms*, Second Edition, Springer-Verlag, New York, 1993.

[Loz11] Álvaro Lozano-Robledo, *Elliptic Curves, Modular Forms and their L-functions*, American Mathematical Society, Princeton, New Jersey, 2011.

[Mur23] Sameer Murthy, *Black holes and modular forms in string theory*, available at https://arxiv.org/abs/2305.11732, 2023.

[Ser73] J-P. Serre, *A course in arithmetic*, Springer-Verlag, New York, 1973.

[Sri19] Shruthi Sridhar, *Modular Forms and Jacobi's Four Square Theorem*, 2019.

[Wil95] Andrew Wiles, *Modular Elliptic Curves and Fermat's Last Theorem.* Ann. of Math. 141 (1995), no. 3, 443-551.