

THESIS

THE MODULAR GROUP AND MODULAR FORMS

Submitted by

Eric Schmidt

Department of Mathematics

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2010

COLORADO STATE UNIVERSITY

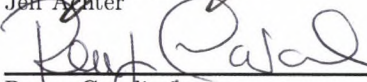
May 17, 2010

WE HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER OUR SUPERVISION BY ERIC SCHMIDT ENTITLED "THE MODULAR GROUP AND MODULAR FORMS" BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE.

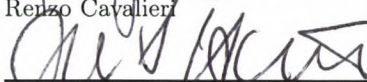
Committee on Graduate Work



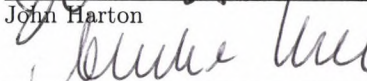
Jeff Achter



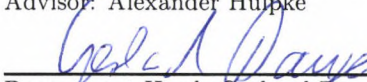
Renzo Cavalieri



John Harton



Advisor: Alexander Hulpke



Department Head: Gerhard Dangelmayr

ABSTRACT OF THESIS

THE MODULAR GROUP AND MODULAR FORMS

We prove some results about the structure of $SL_2(\mathbb{Z})$ and related groups. We define modular forms for this group and develop the basic theory. We then use the theory of lattices to construct examples of modular forms.

Eric Schmidt
Department of Mathematics
Colorado State University
Fort Collins, CO 80538
Summer 2010

1 Introduction

In this paper, we are mainly concerned with the group $SL_2(\mathbb{Z})$ and of the modular forms that may be defined for it. We first prove some results about the structure of $SL_2(\mathbb{Z})$. We then define modular forms for this group, and develop the basic theory. Finally, we use the theory of lattices to construct some examples of modular forms, namely the theta series. The material is derived from the following sources:

- Bruiner, Jan; Geer, Gerard van der; Harder, Günther; Zagier, Don. *The 1-2-3 of Modular Forms* (2008) Springer.
- Elkies, Noam. “Theta functions and weighted theta functions of Euclidean lattices, with some applications” (2009) <http://www.math.harvard.edu/~elkies/aws09.pdf>.
- Koecher, Max; Krieg, Aloys. *Elliptische Funktionen und Modulformen* 2nd ed. (2007) Springer.
- Miyake, Toshitsune. *Modular forms* (1989, 2006) Springer.
- Serre, Jean-Pierre. *A Course in Arithmetic* (1973) Springer-Verlag.

2 The structure of $SL_2(\mathbb{Z})$

2.1 Some relevant groups

If n is a positive integer, the group $GL_n(\mathbb{R})$ is, by definition, the group of all $n \times n$ matrices with real entries that are invertible (i.e., matrices with nonzero determinant). The subgroup $SL_n(\mathbb{R})$ is the group of elements of $GL_n(\mathbb{R})$ with determinant 1. We can also construct the analogous matrix groups over the integers. The group $GL_n(\mathbb{Z})$ consists of the $n \times n$ integer matrices with inverse also an integer matrix, and $SL_n(\mathbb{Z})$ is the subgroup of $GL_n(\mathbb{Z})$ of elements of determinant 1. At present, we will only be concerned with matrices of dimension 2×2 .

The elements

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

of $SL_2(\mathbb{Z})$ are elementary matrices: T acts on a 2×2 matrix from the left by adding the second row to the first row, and S acts by swapping the two rows and then negating the first row.

Convention. When we speak of a matrix $A \in GL_2(\mathbb{R})$, we will denote the entries of A by

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

Theorem 1. *The group $SL_2(\mathbb{Z})$ is generated by T and S .*

Proof. Let Λ be the subgroup of $SL_2(\mathbb{Z})$ generated by T and S . Consider an arbitrary element A of $SL_2(\mathbb{Z})$. We will prove that $A \in \Lambda$. The proof is by induction on the absolute value of c .

First, suppose that $c = 0$. Then, by the determinant condition, $a = d = \pm 1$. Since $S^2 = -I$, and for $n \in \mathbb{Z}$,

$$T^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix},$$

the claim is verified in this case.

Now suppose that $|c| > 0$ and that all elements of $\mathrm{SL}_2(\mathbb{Z})$ whose 2,1 entry has magnitude smaller than $|c|$ are in Λ . There exists an integer n such that $|a + nc| < |c|$. Then, since

$$ST^n A = \begin{pmatrix} -c & -d \\ a + nc & b + nd \end{pmatrix}$$

is in Λ by the inductive hypothesis, we have $A \in \Lambda$. □

2.2 Group actions

Let $\overline{\mathbb{C}}$ denote the complex plane together with a new “point at infinity” ∞ . We will also write $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\} \subseteq \overline{\mathbb{C}}$. The group $\mathrm{GL}_2(\mathbb{R})$ acts from the left on $\overline{\mathbb{C}}$ as follows. For $z \in \mathbb{C}$, write

$$Az = \frac{az + b}{cz + d},$$

where the result is defined to be ∞ if the denominator vanishes. We define the action at ∞ by taking the limit as $z \rightarrow \infty$; that is, we set $A\infty = a/c$. This is a real number if $c \neq 0$ and is ∞ otherwise.

We easily see that, if I is the identity matrix, then $Iz = z$ for $z \in \overline{\mathbb{C}}$. So, to verify that we have a group action, it remains to show that $A(Bz) = (AB)z$ for any $A, B \in \mathrm{SL}_2(\mathbb{Z})$. If we suppose that

$z \neq \infty$ and that denominators do not vanish, we calculate:

$$\begin{aligned}
\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \left[\begin{pmatrix} e & f \\ g & h \end{pmatrix} \cdot z \right] &= \frac{a \frac{ez+f}{gz+h} + b}{c \frac{ez+f}{gz+h} + d} \\
&= \frac{a(ez+f) + b(gz+h)}{c(ez+f) + d(gz+h)} \\
&= \frac{(ae+bg)z + af + bh}{(ce+dg)z + cf + dh} \\
&= \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix} \cdot z \\
&= \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} \right] \cdot z.
\end{aligned}$$

A similar calculation holds if $z = \infty$ or zero denominators occur.

Let H be the set of complex numbers with positive imaginary part. The group action restricted to $\text{SL}_2(\mathbb{Z})$ is an action on H . That is, H is closed under the action. To see this, let $z \in H$. We note first that $cz + d \neq 0$, for otherwise we would have either that z is a real number $-d/c$, or that $c = d = 0$, which violates the determinant condition. We can now show that the operation takes points in H to points in H . More specifically, we can show that, if $z \in H$, then $\text{Im}((az+b)/(cz+d)) = \text{Im}(z)/|cz+d|^2 > 0$. To do this, we write $z = x + yi$, where $x, y \in \mathbb{R}$ and $y > 0$. Then

$$\begin{aligned}
\frac{az+b}{cz+d} &= \frac{a(x+yi)+b}{c(x+yi)+d} \\
&= \frac{ax+b+ayi}{cx+d+cyi} \\
&= \frac{(ax+b+ayi)(cx+d-cyi)}{(cx+d+cyi)(cx+d-cyi)} \\
&= \frac{acx^2+adx+bcx+bd+acy^2+(ad-bc)yi}{|cx+d+cyi|^2} \\
&= \frac{acx^2+adx+bcx+bd+acy^2+yi}{|cz+d|^2}.
\end{aligned}$$

This has imaginary part $y/|cz+d|^2$, which verifies our assertion. Therefore this point is in H .

Applying the definitions, we see that the generators T and S act as $Tz = z + 1$ and $Sz = -1/z$.

The quotient $\mathrm{SL}_2(\mathbb{Z})/Z(\mathrm{SL}_2(\mathbb{Z}))$ of $\mathrm{SL}_2(\mathbb{Z})$ by its center is denoted $\mathrm{PSL}_2(\mathbb{Z})$. The center of $\mathrm{SL}_2(\mathbb{Z})$ in fact consists of just $\pm I$, so $\mathrm{PSL}_2(\mathbb{Z})$ is obtained from $\mathrm{SL}_2(\mathbb{Z})$ by considering matrices equivalent if they differ by a factor of -1 . For any $A \in \mathrm{SL}_2(\mathbb{Z})$, we see that A and $-A$ act identically on H . Thus the action of $\mathrm{SL}_2(\mathbb{Z})$ induces an action of $\mathrm{PSL}_2(\mathbb{Z})$ on H .

2.3 Fundamental domain

Define the subset $F = \{z \in H : -1/2 \leq \mathrm{Re}(z) \leq 1/2, |z| \geq 1\}$ of the upper half plane. This region is bounded by two vertical lines and a circle centered at 0, which intersect at the points $\pm 1/2 + (\sqrt{3}/2)i$, which are also the points $e^{\pi i/3}$ and $e^{2\pi i/3}$. We will write ρ for the point $e^{2\pi i/3}$. Notice that $e^{\pi i/3} = S\rho = -\bar{\rho}$. It is easy to see that if $z \in F$, then $\mathrm{Im}(z) \geq \sqrt{3}/2$, and the inequality is strict if z is in the interior of F .

Theorem 2. *The set F is a fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})$ on H , in the sense that (i) each point of H is equivalent to a point in F under $\mathrm{SL}_2(\mathbb{Z})$, and (ii) no two distinct points in the interior of F are equivalent under $\mathrm{SL}_2(\mathbb{Z})$.*

Proof. To prove the first claim, let $z \in H$. We claim that we can find an element A of $\mathrm{SL}_2(\mathbb{Z})$ such that $|cz + d|$ is minimal. To prove this, choose a real number x such that the set $S = \{(c, d) \in \mathbb{Z}^2 : \gcd(c, d) = 1, |cz + d| < x\}$ is nonempty. We will prove that S is also finite; this will imply that we can take an element of S that minimizes $|cz + d|$ over S , and since c and d are coprime we can find a and b so that the matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has determinant 1.

Now, $|cz + d| \geq |\mathrm{Im}(cz + d)| = |c|\mathrm{Im}(z)$. So, for all sufficiently large c , we have $|cz + d| \geq x$ (regardless of d). Hence, to obtain a member of S , there are only finitely many choices for c . Again, $|cz + d| \geq |\mathrm{Re}(cz + d)| = |\mathrm{Re}(cz) + d|$, so for any fixed c , there are only finitely many choices for d . This shows that S is finite.

We are now justified in taking $A \in \mathrm{SL}_2(\mathbb{Z})$ with $|cz + d|$ minimal. From the formula $\mathrm{Im}(Az) = \mathrm{Im}(z)/|cz + d|^2$, we see that A maximizes $\mathrm{Im}(Az)$. Since $T^n Az = Az + n$ for any integer n , we can choose n so that $-1/2 \leq \mathrm{Re}(T^n Az) \leq 1/2$. The condition $|T^n Az| \geq 1$ also holds. For, if not, then from $|T^n Az| < 1$ we deduce $\mathrm{Im}(S(T^n Az)) = \mathrm{Im}(T^n Az)/|T^n Az|^2 > \mathrm{Im}(T^n Az) = \mathrm{Im}(Az)$, contradicting the maximality of $\mathrm{Im}(Az)$. We conclude that $T^n Az \in F$.

To prove the second claim, take two points z_1, z_2 in the interior of F . We can assume without loss of generality that $\text{Im}(z_2) \geq \text{Im}(z_1)$. Suppose that $Az_1 = z_2$. Then, since $\text{Im}(z_2) = \text{Im}(z_1)/|cz_1 + d|^2$, we have $|cz_1 + d| \leq 1$. Since also $|cz_1 + d| \geq |\text{Im}(cz_1 + d)| = |c|\text{Im}(z) > |c|\sqrt{3}/2 > |c|/2$, we must have $|c| \leq 1$. Now, $|\text{Im}(z_1)| = |\text{Im}(\pm z_1 + d)|$, but $|\text{Re}(z_1)| \leq |\text{Re}(\pm z_1 + d)|$ (for, if, $d \neq 0$, then $1/2 \leq |\text{Re}(\pm z_1 + d)|$). It follows from this that $|\pm z_1 + d| \geq |z_1| > 1$. This means we cannot have $c = \pm 1$. The only remaining possibility is $c = 0$. Then, $b = d = \pm 1$, and A is of the form $\pm T^n$. Since the real parts of z_1 and z_2 differ by less than 1, it must be that $n = 0$, and so $A = \pm I$ and $z_1 = z_2$. \square

2.4 Classification of elements of $\text{GL}_2(\mathbb{R})$

The center of $\text{GL}_2(\mathbb{R})$ consists of the matrices xI with x a nonzero real number.

Definition 1. If $A \in \text{GL}_2(\mathbb{R})$ is not central, then it is called *elliptic* if $(\text{tr } A)^2 < 4 \det A$, *parabolic* if $(\text{tr } A)^2 = 4 \det A$, and *hyperbolic* if $(\text{tr } A)^2 > 4 \det A$.

The elements of $\text{GL}_2(\mathbb{R})$ can be described in terms of the fixed points of their actions on $\overline{\mathbb{C}}$. The central matrices fix every point. The other elements are characterized by the following theorem.

Theorem 3. *A matrix $A \in \text{GL}_2(\mathbb{R})$ is elliptic if and only if it has two distinct, complex conjugate, fixed points (and no others); A is parabolic if and only if it has one fixed point in $\overline{\mathbb{R}}$ (and no others); and A is hyperbolic if and only if it has two distinct fixed points in $\overline{\mathbb{R}}$ (and no others).*

Proof. To prove this, consider a noncentral matrix $A \in \text{GL}_2(\mathbb{R})$. First, suppose that $c \neq 0$. Then ∞ is not a fixed point. Therefore, the fixed points are the solutions of the equation $(az + b)/(cz + d) = z$, or $cz^2 + (d - a)z - b = 0$. The discriminant of this equation is $(d - a)^2 + 4bc = d^2 - 2ad + a^2 + 4bc = (a + d)^2 - 4(ad - bc) = (\text{tr } A)^2 - 4 \det A$. Hence, if A is elliptic, there are two distinct, complex conjugate fixed points; if A is parabolic there is one, real, fixed point; and if A is hyperbolic, then there are two distinct, real, fixed points.

Next, suppose $c = 0$. Then ∞ is a fixed point. Any other fixed points are the solutions of the equation $(az + b)/d = z$, or $(d - a)z - b = 0$. If $a = d$, then there are no solutions (for if $b = 0$ then $A = aI = dI$ is central). Thus, there is only one fixed point, ∞ . If, however $a \neq d$, then $z = b/(d - a)$ is a real fixed point. Moreover, since $(\text{tr } A)^2 - 4 \det A = (a + d)^2 - 4ad = a^2 + 2ad + d^2 - 4ad =$

$a^2 - 2ad - d^2 = (a - d)^2$, the matrix A is parabolic if $a = d$, but hyperbolic if $a \neq d$. These results confirm the claim when $c = 0$. \square

We pause to give some examples of elliptic, parabolic, and hyperbolic matrices and the fixed points of their action on $\overline{\mathbb{C}}$.

1. Elliptic elements. The matrix S is elliptic, with trace 0 and determinant 1. The fixed points of its action are $\pm i$. Another elliptic matrix is $\begin{pmatrix} 3 & -2 \\ 5 & 2 \end{pmatrix}$, with trace 5 and determinant 16. Its fixed points are $(1 \pm \sqrt{-39})/10$.
2. Parabolic elements. The matrix T is parabolic, with trace 2 and determinant 1. As it induces the action $z \mapsto z + 1$, it has no fixed points in \mathbb{C} , but ∞ is a fixed point. Another parabolic matrix is $\begin{pmatrix} 1 & -2 \\ 8 & -7 \end{pmatrix}$, with trace -6 and determinant 9. Its fixed point is $1/2$.
3. Hyperbolic elements. Any element of $\text{GL}_2(\mathbb{R})$ with negative determinant is hyperbolic. Another example is $\begin{pmatrix} 4 & 1 \\ 1 & 5 \end{pmatrix}$, with trace 9 and determinant 19. Its fixed points are $(-1 \pm \sqrt{5})/2$. Another example is $\begin{pmatrix} -2 & 1 \\ 0 & -1 \end{pmatrix}$, with trace -3 and determinant 2. Its fixed points are ∞ and 1.

We now concentrate on $\text{SL}_2(\mathbb{R})$. In this subgroup, matrices A other than $\pm I$ are elliptic, parabolic, or hyperbolic, as $|\text{tr } A| < 2$, $|\text{tr } A| = 2$, or $|\text{tr } A| > 2$. We can also characterize the classes of matrices in terms of their eigenvalues. The characteristic polynomial of A is $x^2 - (\text{tr } A)x + 1$, which has discriminant $(\text{tr } A)^2 - 4$. We thus easily deduce that a matrix is elliptic if and only if it has no real eigenvalues; it is parabolic if and only if it has a double eigenvalue of either 1 or -1 ; and it is hyperbolic if and only if it has two distinct real eigenvalues.

Theorem 4. *In $\text{SL}_2(\mathbb{R})$, elements of finite order must be central or elliptic.*

Proof. Suppose that A is parabolic. Then the Jordan canonical form of A has diagonal entries equal to λ , the eigenvalue of A . Since A is not central, the entry in the upper-right corner must be nonzero (and thus equal to 1). Then, the n th power of the Jordan canonical form has the entry in the upper-right corner equal to $\lambda^{n-1}n$. It follows that A has infinite order. Next, suppose that A is hyperbolic. Then the Jordan canonical form of A is a diagonal matrix with two distinct real numbers on the diagonal. Taking the n th power of this matrix replaces the diagonal entries with

their n th power. The only real roots of unity are 1 and -1 , but these cannot be the diagonal entries, for then the matrix would have determinant -1 . Thus, in this case too, A has infinite order. \square

2.5 Elements of $\mathrm{SL}_2(\mathbb{Z})$ of finite order

We now again restrict our attention, this time to $\mathrm{SL}_2(\mathbb{Z})$. We will see that if an element of $\mathrm{SL}_2(\mathbb{Z})$ has finite order, there are only a few possibilities for what that order is.

Theorem 5. *In $\mathrm{SL}_2(\mathbb{Z})$, the element I has order 1, the element $-I$ has order 2, and all other elements of finite order are elliptic with order 3, 4, or 6.*

Proof. The first two assertions are obvious. Let A be a noncentral element of $\mathrm{SL}_2(\mathbb{Z})$ of finite order. We already know that A must be elliptic. Thus, it has no real eigenvalues, and so the characteristic polynomial of A is irreducible over \mathbb{Z} . Therefore the characteristic polynomial of A is its minimal polynomial, and so the minimal polynomial has degree 2. Since A has finite order, it satisfies the polynomial $x^n - 1$ for some $n > 1$. The irreducible factors of $x^n - 1$ are precisely the cyclotomic polynomials $\Phi_d(x)$ for the divisors d of n . Therefore the minimal polynomial of A is $\Phi_k(x)$ for some k . By the same factorization theorem just cited, we see that $\Phi_k(x)$ divides $x^k - 1$, but does not divide any $x^m - 1$ with $m < k$. So, A has order k . But $\Phi_k(x)$ has degree $\phi(k)$ (where ϕ denotes Euler's totient function) and also has degree 2, so $\phi(k) = 2$. The only values of k for which $\phi(k) = 2$ are 3, 4, and 6. (This may be verified using the product formula for ϕ . From this formula we see that if p is a prime, a is a positive integer, and p^a divides k , then both $p - 1$ and p^{a-1} divide $\phi(k)$. This restricts the possible prime factors of k to 2, which can occur at most twice, and 3, which can occur at most once.) \square

Soon we will see another way to prove this theorem.

We know that every elliptic element of $\mathrm{SL}_2(\mathbb{Z})$ fixes exactly one point of H . Conversely, for a point of H , we can ask what its stabilizer in $\mathrm{SL}_2(\mathbb{Z})$ is. The stabilizer must include the central matrices $\pm I$, and any other elements of the stabilizer must be elliptic. We will now show that for most points, the stabilizer is just $\{\pm I\}$.

Theorem 6. *Let z be a point in the fundamental domain D , and let A be an elliptic element of $\mathrm{SL}_2(\mathbb{Z})$ that fixes z . Then z is either i , ρ , or $-\bar{\rho}$. Moreover the stabilizer in $\mathrm{SL}_2(\mathbb{Z})$ of i is the cyclic*

group of order 4 generated by S , that of ρ the cyclic group of order 6 generated by ST , and that of $-\bar{\rho}$ the cyclic group of order 6 generated by TS .

Proof. Consider an element A of $\mathrm{SL}_2(\mathbb{Z})$ that fixes $z \in D$. By the proof of Theorem 3, $c \neq 0$ and $cz^2 + (d - a)z - b = 0$. So, applying the quadratic formula, we obtain $\mathrm{Re}(z) = (a - d)/(2c)$. Since $Az = z$, the equation $\mathrm{Im}(Az) = \mathrm{Im}(z)/|cz + d|^2$ implies that $|cz + d| = 1$. Since also $|cz + d| \geq |c| |\mathrm{Im}(z)| > |c|/2$, we have $c = \pm 1$. If $c = -1$ we can negate all entries of A and obtain an equivalent action, so suppose without loss of generality that $c = 1$. Then the real part of z is either an integer or half-integer, and so is either 0, $1/2$, or $-1/2$. Next, since $|\mathrm{Im}(z)| = |\mathrm{Im}(z + d)|$, but $|\mathrm{Re}(z)| \leq 1/2 \leq |\mathrm{Re}(z + d)|$ (when $d \neq 0$), we have $1 = |z + d| \geq |z|$. Since by definition $|z| \geq 1$, we have $|z| = 1$. The only such points in H with real part 0, $1/2$, or $-1/2$ are i , ρ and $-\bar{\rho}$ respectively.

It is now straightforward to determine the stabilizers of these 3 points, using the facts $|z + d| = 1$ and $\mathrm{Re}(z) = (a - d)/2$. If we continue to assume $c = 1$, then for the point $z = i$, the only value of d such that $|z + d| = 1$ is 0. Since $\mathrm{Re}(z) = 0$, we then conclude that $a = 0$. Finally, the determinant condition shows that $b = -1$. Thus, we have the matrix S . If we had used $c = -1$ we would have obtained $-S$. These two matrices together with $\pm I$ are the cyclic group generated by S .

For the point $z = \rho$, the only possible values for d with $|z + d| = 1$ are 0 and 1. Since $\mathrm{Re}(z) = -1/2$, we have $a = d - 1$, so for a we get the values -1 and 0, respectively, and from the determinant condition we get $b = -1$ in either case. The two matrices, together with their negatives and $\pm I$, constitute the cyclic group generated by ST .

The analysis for $-\bar{\rho}$ is similar. □

Corollary 1. *An element of $\mathrm{SL}_2(\mathbb{Z})$ has finite order if and only if it is central or elliptic. An elliptic element is conjugate to an element fixing i or ρ .*

Proof. We have shown that any point in H is equivalent by the action of $\mathrm{SL}_2(\mathbb{Z})$ to a point of D . Moreover, ρ and $-\bar{\rho}$ are equivalent by the matrix S . Since any elliptic element A fixes a point in H , it is conjugate to an element fixing a point $z \in D$, which must be i , ρ , or $-\bar{\rho}$ since A is noncentral. This implies the second assertion. For the first assertion, we already know that elements of finite order are central or elliptic. Conversely, all central or elliptic elements are conjugate to elements fixing a point in D , and we have seen that all such elements have finite order. □

Notice that we have also obtained another proof of Theorem 5, for, since any noncentral element of finite order is elliptic, it lies in the stabilizer of a point in the upper half plane, and all such stabilizers have order 2, 4, or 6. Moreover, we see that all the element orders specified in Theorem 5 actually occur.

2.6 $\mathrm{PSL}_2(\mathbb{Z})$ as a free product

The *free product* of two groups is, essentially, the group generated by the two groups with no relations between the elements of the groups. More precisely, we have the following definition.

Definition 2. Let G and H be groups with presentations $\{B \mid R\}$ and $\{C \mid S\}$ respectively, where B and C are disjoint. The *free product* of G and H is the group with presentation $\{B \cup C \mid R \cup S\}$.

(In order for this to be a proper definition, it is necessary to show that the free product does not depend on the particular presentations of G and H chosen. We omit the proof.)

From Theorem 1, we deduce that the matrices S and ST (which generate the stabilizers of i and ρ respectively) generate the group $\mathrm{SL}_2(\mathbb{Z})$. Thus $\mathrm{SL}_2(\mathbb{Z})$ is generated by an element of order 4 and an element of order 6. Since $S^2 = (ST)^3 = -I$, the images of these points in $\mathrm{PSL}_2(\mathbb{Z})$ have orders 2 and 3. It can be shown that $\mathrm{PSL}_2(\mathbb{Z})$ is in fact the free product of the cyclic groups of order 2 and 3 generated by the images of S and T in $\mathrm{PSL}_2(\mathbb{Z})$. In other words, $\mathrm{PSL}_2(\mathbb{Z})$ is abstractly the group presented by $\{x, y \mid x^2 = y^3 = 1\}$ via an isomorphism sending S to x and ST to y . A consequence of this is that any group that can be generated by an element of order 2 and an element of order 3 can be obtained as a quotient of $\mathrm{PSL}_2(\mathbb{Z})$. This is the case for most families of nonabelian finite simple groups.

3 Modular forms

3.1 Definition of modular forms

Definition 3. Let k be an integer. A meromorphic function $f : H \rightarrow \mathbb{C}$ is called a *weakly modular function* of weight k [for $\mathrm{SL}_2(\mathbb{Z})$] if for any $A \in \mathrm{SL}_2(\mathbb{Z})$ and $z \in H$, the relation $f(Az) = (cz+d)^k f(z)$ holds.

We will call the relation occurring in the definition the *transformation rule* for modular forms. If we let $A = T$ in the definition, then the transformation rule becomes $f(z+1) = f(z)$. Thus, if f is weakly modular, the value of f at z depends only on $q = e^{2\pi iz}$, since from q we can recover z up to an integer difference. Define the function \tilde{f} by $\tilde{f}(q) = f(e^{2\pi iz})$. Now, if we let $\mathrm{Im}(z) \rightarrow \infty$, then $q \rightarrow 0$. Thus, we say that f is meromorphic at ∞ if \tilde{f} is meromorphic at 0; that is, if we can write

$$\tilde{f}(q) = \sum_{n \in \mathbb{Z}} a_n q^n.$$

with $a_n = 0$ for all sufficiently small n . Such a series is called a Fourier series for f . Similarly, f is holomorphic at ∞ precisely when \tilde{f} is holomorphic at 0, or, equivalently, if we may take $a_n = 0$ for all $n < 0$. If f is holomorphic at ∞ , then we define the value of f at ∞ to be the value of the Fourier expansion when $q = 0$; that is, $f(\infty) = a_0$.

Definition 4. Let k be an integer. A function $f : H \rightarrow \mathbb{C}$ is called a *modular form* of weight k [for $\mathrm{SL}_2(\mathbb{Z})$] if it is weakly modular and is holomorphic everywhere, including at ∞ .

(In the above definitions we have specified “for $\mathrm{SL}_2(\mathbb{Z})$ ” in brackets because these concepts can be defined with reference to many other groups as well.)

Notice that if k is odd, only the zero function can be a modular form of weight k . This is true because if we use the matrix $-I$ in the transformation rule, we see that a modular form is required to satisfy $f(z) = (-1)^k f(z) = -f(z)$ if k is odd, and this implies that $f(z) = 0$. For this reason we will henceforth assume that k is even.

When showing that a certain function is a modular form, it can be cumbersome to verify the transformation rule for every $A \in \mathrm{SL}_2(\mathbb{Z})$. Fortunately, this is not necessary. To demonstrate this, it will be convenient to introduce the notation $j(A, z) = cz + d$. With this definition, the transformation rule for modular forms becomes $f(Az) = j(A, z)^k f(z)$. Take $A, B \in \mathrm{SL}_2(\mathbb{Z})$. We claim that $j(AB, z) = j(A, Bz)j(B, z)$. To prove this, write

$$B = \begin{pmatrix} e & f \\ g & h \end{pmatrix}.$$

Then, we compute,

$$\begin{aligned} j(AB, z) &= (ce + dg)z + cf + dh \\ &= c(ez + f) + d(gz + h) \\ &= \left(c \frac{ez + f}{gz + h} + d \right) (gz + h) \\ &= j(A, Bz)j(B, z). \end{aligned}$$

This identity implies that if both A and B satisfy the transformation rule, then so do AB and A^{-1} . By induction, we deduce that if the transformation rule is satisfied by a generating set for $\mathrm{SL}_2(\mathbb{Z})$, such as the matrices T and S , then it is satisfied by all matrices in $\mathrm{SL}_2(\mathbb{Z})$. We will apply this later.

3.2 Modular forms as a graded algebra

Let M_k be the set of modular forms of weight k . This set is a vector space over \mathbb{C} under the pointwise sum and scalar product operations. In particular, the sum and scalar multiple of modular forms of weight k is also of weight k . This is straightforward to verify. For instance, assuming $f, g \in M_k$,

and $A \in \mathrm{SL}_2(\mathbb{Z})$, then we find that $(f + g)(Az) = f(Az) + g(Az) = (cz + d)^k f(z) + (cz + d)^k g(z) = (cz + d)^k (f + g)(z)$. Note also that the sum and scalar products of holomorphic functions are again holomorphic (even at ∞).

We can also consider the pointwise product of functions. If $f \in M_{k_1}$ and $g \in M_{k_2}$ then $(fg)(Az) = f(Az)g(Az) = (cz + d)^{k_1} f(z)(cz + d)^{k_2} g(z) = (cz + d)^{k_1 + k_2} (fg)(z)$. Thus $fg \in M_{k_1 + k_2}$. Therefore, if we form the direct sum

$$\mathcal{M} = \bigoplus_{k \in 2\mathbb{Z}} M_k,$$

then \mathcal{M} is closed under multiplication. Since each space M_k is closed under addition and scalar multiplication, \mathcal{M} forms an algebra over \mathbb{C} . Moreover, the fact that the weight of a product of modular forms is the sum of the weights means precisely that \mathcal{M} is a *graded* algebra.

3.3 Eisenstein series

Definition 5. For even $k > 2$ and $z \in H$, the *Eisenstein series* of index k is

$$G_k(z) = \sum_{\substack{m, n \in \mathbb{Z} \\ (m, n) \neq (0, 0)}} \frac{1}{(mz + n)^k}.$$

We are going to show that the Eisenstein series are modular forms, which will require a bit of work.

Lemma 1. *Let K be a compact subset of H . There exists a positive real number γ such that for all $m, n \in \mathbb{R}$ and all $z \in K$,*

$$\gamma |mi + n| \leq |mz + n|.$$

Proof. If $m = n = 0$, then the result is true regardless of the values of γ or z . So we may assume that either m or n is nonzero. By rescaling, we may then suppose that (m, n) lies on the unit circle C in \mathbb{R}^2 . Then the desired conclusion becomes

$$\gamma \leq |mz + n|.$$

Consider the function on $K \times C$ given by $(z, m, n) \mapsto |mz + n|$. Since $z \in H$, this function never vanishes (see the argument on page 3). Since $K \times C$ is compact, the function takes on a minimum value γ , which by the preceding remark is positive. \square

Lemma 2. *The preceding result holds even if we replace K by the fundamental domain D .*

Proof. Let $z \in D$ and $m, n \in \mathbb{R}$. As in the previous lemma, we may suppose that $m^2 + n^2 = 1$. If $|m| > 1/2$ we obtain $|mz + n| \geq |\operatorname{Im}(mz + n)| \geq \sqrt{3}|m|/2 > \sqrt{3}/4$. If $|m| \leq 1/2$ then $|\operatorname{Re}(mz)| \leq 1/4$ and also $|n| > 1/2$, so that we obtain $|mz + n| \geq |\operatorname{Re}(mz + n)| \geq 1/4$. Therefore, we may take $\gamma = 1/4$. \square

Theorem 7. *For even $k > 2$, the Eisenstein series G_k converges absolutely and is holomorphic on H .*

Proof. It suffices to show that $G_k(i)$ converges absolutely. For then Lemma 1 will show that $G_k(z)$ converges absolutely for any $z \in H$ and (by the Weierstrass M test) converges uniformly in any compact subset of H . The uniform convergence implies that G_k is holomorphic.

Now, to show that $G_k(i)$ converges absolutely, we need to show that

$$\sum_{\substack{m, n \in \mathbb{Z} \\ (m, n) \neq (0, 0)}} \frac{1}{|mi + n|^k} = \sum_{\substack{m, n \in \mathbb{Z} \\ (m, n) \neq (0, 0)}} (m^2 + n^2)^{-k/2} \quad (*)$$

converges. The terms in which $n = 0$ sum to

$$2 \sum_{m=1}^{\infty} m^{-k} = 2\zeta(k).$$

Likewise, the terms in which $m = 0$ also sum to $2\zeta(k)$. Notice that for any $m, n \in \mathbb{R}$, $m^2 + n^2 \geq |mn|$. This follows from the observation that if $m, n \geq 0$, then $m^2 + n^2 - mn \geq m^2 + n^2 - 2mn = (m - n)^2 \geq 0$. Therefore the sum of the remaining terms in the series (those in which neither m nor n is 0) is bounded above by

$$\sum_{m, n \in \mathbb{Z} \setminus \{0\}} |mn|^{-k/2}.$$

The sum of the terms in the latter series for which $m, n > 0$ is bounded above by

$$\left(\sum_{m=1}^{\infty} m^{-k/2} \right)^2 = \zeta^2(k/2).$$

By multiplying the latter constant by 4, we account for all possible signs of m and n .

We have thus shown that the series (*) is bounded above by $4(\zeta(k) + \zeta^2(k/2))$, and thus converges. \square

Theorem 8. *The Eisenstein series G_k is a modular form of weight k . Moreover, $G_k(\infty) = 2\zeta(k)$.*

Proof. Any member of $SL_2(\mathbb{Z})$ acts on \mathbb{Z}^2 as an invertible linear transformation and so yields a permutation of $\mathbb{Z}^2 \setminus \{(0, 0)\}$. Take $A \in SL_2(\mathbb{Z})$. To verify the transformation rule for modular forms, we calculate

$$\begin{aligned} G_k(Az) &= \sum_{\substack{m, n \in \mathbb{Z} \\ (m, n) \neq (0, 0)}} \frac{1}{\left(m \frac{az+b}{cz+d} + n\right)^k} \\ &= (cz+d)^k \sum_{\substack{m, n \in \mathbb{Z} \\ (m, n) \neq (0, 0)}} \frac{1}{(m(az+b) + n(cz+d))^k} \\ &= (cz+d)^k \sum_{\substack{m, n \in \mathbb{Z} \\ (m, n) \neq (0, 0)}} \frac{1}{((am+cn)z + (bm+dn))^k} \\ &= (cz+d)^k \sum_{\substack{m, n \in \mathbb{Z} \\ (m, n) \neq (0, 0)}} \frac{1}{(m'z + n')^k} \end{aligned}$$

where $(m', n') = A^T(m, n)$. Since A^T simply permutes the indices, the last expression equals $(cz+d)^k G_k(z)$. Thus G_k satisfies the transformation rule. We have already shown that G_k is holomorphic on H , so all that is left to do is show that G_k is holomorphic at ∞ and determine $G_k(\infty)$. To do this, we take the limit as $\text{Im}(z) \rightarrow \infty$. We need only consider $\text{Im}(z) \geq 1$. Therefore, by applying a suitable translation T^n , we may assume that $z \in D$. (Since G_k satisfies the transformation rule for modular forms, we know that $G_k(T^n z) = G_k(z)$.) Now Lemma 2 implies that G_k converges uniformly in D , so we may compute the limit of the series term by term. For $m \neq 0$, we have

$1/(mz + n)^k \rightarrow 0$ as $\text{Im}(z) \rightarrow \infty$, and then the remaining terms sum to

$$2 \sum_{n=1}^{\infty} n^{-k} = 2\zeta(k).$$

Thus G_k is holomorphic at ∞ and has the claimed value. □

3.4 Cusp forms

Definition 6. A modular form f is called a *cusp form* if $f(\infty) = 0$.

For an example, first define

$$g_4 = 60G_4 \quad g_6 = 140G_6.$$

Then, define the *discriminant*:

$$\Delta = g_4^3 - 27g_6^2.$$

Since both g_4^3 and g_6^2 have weight 12, this is a modular form of weight 12. Moreover, the coefficients have been chosen so that Δ vanishes at ∞ . This can be verified by noting that by Theorem 9, $G_4(\infty) = 2\zeta(4)$ and $G_6(\infty) = 2\zeta(6)$, and by using the values

$$\zeta(4) = \frac{\pi^4}{90} \quad \zeta(6) = \frac{\pi^6}{945}.$$

Thus, Δ is a cusp form of weight 12. However, we do not yet know whether Δ is just the zero function. We will soon show that Δ does not vanish anywhere in H .

The set of cusp forms of weight k is denoted S_k . This is a vector subspace of M_k . If $M_k \neq 0$, then the codimension of S_k in M_k is 1. This holds since if $f, g \in M_k$ but are not cusp forms, then for a suitable $v \in \mathbb{C}$, the modular form $f + vg$ vanishes at ∞ , so f and g are linearly dependent in the quotient M_k/S_k . Since the Eisenstein series are not cusp forms, for even $k > 2$ we can write M_k as a direct sum $S_k \oplus \mathbb{C}G_k$, where $\mathbb{C}G_k$ is the 1-dimensional subspace of M_k spanned by G_k .

3.5 Zeros of modular forms

If f is a meromorphic function on H and $w \in H$, then for z near w we can write

$$f(z) = \sum_{n \in \mathbb{Z}} a_n (z - w)^n.$$

If $f \neq 0$, then define $v_w(f)$, the order of f at w , to be the smallest n for which $a_n \neq 0$. (Such an n exists because f is meromorphic.) Thus, $v_w(f) \geq 0$ if and only if f is holomorphic at w , and $v_w(f) > 0$ if and only if $f(w) = 0$. We say that f has a *simple zero* at w if $v_w(f) = 1$, and a *simple pole* at w if $v_w(f) = -1$. Notice that for f, g meromorphic, we have $v_w(fg) = v_w(f) + v_w(g)$, as can be seen by multiplying the Laurent series for f and g . If f is weakly modular and meromorphic at ∞ , we also define $v_\infty(f)$ to be the smallest n such that $a_n \neq 0$ in the Fourier expansion described before Definition 4. Now, for $f \in M_k$ and $A \in \mathrm{SL}_2(\mathbb{Z})$, by definition $f(Aw) = (cw + d)^k f(w)$. Since $(cw + d)^k \neq 0$, this means that $v_{Az}(f) = v_z(f)$. Thus, the order of f at w depends only on the orbit of w under the action of $\mathrm{SL}_2(\mathbb{Z})$.

We can show that, if we only count one point from each orbit, a nonzero modular form f has finitely many zeros. More precisely, let \mathcal{O} be the set of orbits of the action of $\mathrm{SL}_2(\mathbb{Z})$ on H . Then we have the following result.

Theorem 9. *Let f be a nonzero modular form of weight k . Then,*

$$v_\infty(f) + \frac{v_i(f)}{2} + \frac{v_\rho(f)}{3} + \sum_{q \in \mathcal{O} \setminus \{i, \rho\}} v_q(f) = \frac{k}{12}.$$

We omit the proof of this theorem; see [Serre] for a proof.

Corollary 2.

- (i) *The Eisenstein series G_4 has a simple zero at ρ , and $v_z(G_4) = 0$ for z not in the orbit of ρ .*
- (ii) *The Eisenstein series G_6 has a simple zero at i , and $v_z(G_6) = 0$ for z not in the orbit of i .*
- (iii) *The discriminant Δ has a simple zero at ∞ and does not vanish in H .*

Proof. These results follow easily from the previous theorem by exploiting the fact that the order of a point is a nonnegative integer. Note that we have shown that G_4 and G_6 are nonzero (they

do not vanish at ∞), so we may apply the theorem to them. For (i), G_4 has weight 4, so the sum occurring in the previous theorem is $4/12 = 1/3$. We find that the only way to achieve this is by setting $v_\rho(G_4) = 1$ and $v_z(G_4) = 0$ for points z inequivalent to ρ . For (ii), G_6 has weight 6, so the sum occurring in the previous theorem is $1/2$. Again, this forces $v_i(G_6) = 1$ and $v_z(G_6) = 0$ for points z inequivalent to i . For (iii), we have just shown that G_6 vanishes at i , but G_4 does not. Therefore, Δ does not vanish at i , and the previous theorem is applicable. Now Δ is of weight 12, so the sum in the theorem is 1. Since Δ does vanish at ∞ , this means that $v_\infty(\Delta) = 1$ and Δ does not vanish elsewhere. \square

3.6 The dimension of M_k

For each even k , we know that M_k is a complex vector space, but we do not yet know how large M_k is, or even whether M_k is finite dimensional. We can use Theorem 9 to give a precise answer. We will need the following:

Theorem 10. *Let k be an even integer. The map $f \mapsto \Delta f$ is a vector space isomorphism from M_k to S_{k+12} .*

Proof. Denote by ϕ the map $f \mapsto \Delta f$. This map is a linear transformation and sends M_k into S_{k+12} . Moreover, if $g \in S_{k+12}$, then g/Δ is weakly modular of weight k and is meromorphic at ∞ . As Δ does not vanish on H , the function g/Δ is holomorphic on H . Moreover, since $v_\infty(g) > 0$ and $v_\infty(\Delta) = 1$, we have that $v_\infty(g/\Delta) \geq 0$, so g/Δ is holomorphic at ∞ as well. Thus g is a modular form of weight k . So the map $g \mapsto g/\Delta$ sends S_{k+12} into M_k , and it is an inverse to ϕ , so ϕ is an isomorphism. \square

Now we can determine the dimension of M_k for all k . Recall that if M_k is nonzero then S_k is of codimension 1 in M_k .

Theorem 11. *Let k be an even integer. If $k < 0$, then $\dim M_k = 0$. Otherwise, if $k \equiv 2 \pmod{12}$, then $\dim M_k = \lfloor k/12 \rfloor$. In all other cases $\dim M_k = \lfloor k/12 \rfloor + 1$.*

Proof. First, suppose $k < 0$. Since the sum in Theorem 9 is nonnegative, we cannot have any nonzero modular forms of weight k , so that $\dim M_k = 0$. From this and the previous theorem we see that

for $0 \leq k \leq 10$, we have $\dim S_k = 0$, so that $\dim M_k \leq 1$. Since the identity function is a nonzero modular form of weight 0, this means that $\dim M_0 = 1$. If $k = 2$, then the sum in Theorem 9 is $1/6$, which is plainly impossible, so $\dim M_2 = 0$. For $k \geq 4$ the modular form G_k is of weight k and is nonzero, so $\dim M_k = 1$ for $4 \leq k \leq 10$.

We have now verified the theorem for $k < 12$. We can now use induction to prove the theorem for $k \geq 12$. Since $\lfloor (k-12)/12 \rfloor = \lfloor k-12 \rfloor - 1$, we just need to show that $\dim M_k = \dim M_{k-12} + 1$. This follows immediately from the fact that M_{k-12} is isomorphic to S_k .

□

This theorem lets us prove the following.

Theorem 12. *The graded \mathbb{C} -algebra \mathcal{M} of all modular forms is generated by G_4 and G_6 .*

Proof. Let \mathcal{N} be the subalgebra of \mathcal{M} generated by G_4 and G_6 . We get $M_0 \subseteq \mathcal{N}$ for free since this space is the image of \mathbb{C} in \mathcal{M} . The space M_2 is 0-dimensional. Since M_4 is 1-dimensional, it is generated by G_4 . For even $k \geq 6$, we use induction to show that $M_k \subseteq \mathcal{N}$. Let $f \in M_k$. As k is even, either k or $k - 6$ is divisible by 4. Thus we can find nonnegative integers α and β such that $k = 4\alpha + 6\beta$. Then $G_4^\alpha G_6^\beta$ is of weight k . Since the Eisenstein series do not vanish at ∞ , we can find a scalar λ such that $g = f - \lambda G_4^\alpha G_6^\beta$ is a cusp form. By Theorem 10, $g = \Delta h$ for some $h \in M_{k-12}$. By the inductive hypothesis, $h \in \mathcal{N}$, and since $\Delta \in \mathcal{N}$ by definition of Δ , we conclude that $f \in \mathcal{N}$.

□

3.7 The j -function

The j -function is defined to be $j = 1728g_4^3/\Delta$. It is a weakly modular function of weight 0. Since Δ does not vanish in H , the j -function is holomorphic there. However, since G_4 is nonzero at ∞ , but Δ has a simple zero at ∞ , the j -function has a simple pole at ∞ , which prevents j from being a modular form.

We can express the j -function as a Fourier series:

$$j(z) = \frac{1}{q} + 744 + 196884q + 21493760q^2 + \cdots,$$

where we recall that $q = e^{2\pi iz}$. The coefficients of this series are all integers. They are connected with the “Monster”, the largest of the sporadic simple groups. We briefly describe this connection, which concerns the irreducible complex representations of the Monster. In general, a complex representation of a group G is a complex vector space V together with a homomorphism ϕ of G into $\text{GL}(V)$, the group of invertible linear transformations of V . We will assume that V is of finite dimension, say, n ; thus, we may take $\text{GL}(V)$ to be the group $\text{GL}_n(\mathbb{C})$ of $n \times n$ invertible complex matrices. We call n the *degree* of the representation. A *subrepresentation* of (V, ϕ) is a subspace W of V closed under $\phi(G)$, and the action of G on W is given by restricting the action on V . If $\phi_1 : G \rightarrow V_1$ and $\phi_2 : G \rightarrow V_2$ are two representations of G , then we can define another representation, the direct sum $\psi = \phi_1 \oplus \phi_2 : G \rightarrow V_1 \oplus V_2$ by $\psi(g) = \phi_1(g) \oplus \phi_2(g)$. The matrix of $\psi(g)$, with respect to the basis of $V_1 \oplus V_2$ obtained from bases of V_1 and V_2 , is then the block diagonal matrix $\begin{pmatrix} \phi_1(g) & 0 \\ 0 & \phi_2(g) \end{pmatrix}$. A representation is *indecomposable* if it is not the direct sum of two representations of smaller degree. For complex representations of finite groups, the concepts of irreducible and indecomposable coincide. Moreover, a finite group has, up to isomorphism, finitely many irreducible complex representations, and it is a general problem to determine what they are, and in particular to determine the degrees of the irreducible representations.

The two smallest irreducible representations of the Monster have degree 1 and 196883, and it is observed that $1 + 196883$ is the coefficient of the q term in the Fourier series in the j -function. In general, each coefficient of the j -function (except for the constant term 744) is a integer linear combination of the irreducible representation degrees of the Monster, with “small” coefficients. (The last proviso is of course what makes the statement interesting; without it, we could merely observe that that any integer is a multiple of 1!)

4 Lattices and theta series

4.1 Lattices

A subset L of the vector space \mathbb{R}^n is an n -dimensional *lattice* if there is a basis $\Omega = \omega_1, \dots, \omega_n$ of \mathbb{R}^n , such that L is the set of all \mathbb{Z} -linear combinations of $\omega_1, \dots, \omega_n$. Let $\langle v_1, v_2 \rangle$ denote the usual inner product of v_1 and v_2 in \mathbb{R}^n . Then the *Gram matrix* of the basis is the $n \times n$ matrix with i, j entry equal to $\langle \omega_i, \omega_j \rangle$. We also call this matrix the Gram matrix of L . However, two different bases of \mathbb{R}^n can determine the same lattice, so a lattice does not have a uniquely determined Gram matrix.

If $\Omega = \omega_1, \dots, \omega_n$ and $\Gamma = \gamma_1, \dots, \gamma_n$ are two bases of \mathbb{R}^n , and A is the Gram matrix of Ω , then we can determine the Gram matrix of Γ as follows. Let B be the change of basis matrix from Γ to Ω . Then the Gram matrix of Γ is $B^T A B$. Verifying this is an exercise in matrix multiplication and applying the bilinearity of the inner product. A particular case occurs when Ω is the standard basis. Then B is the matrix whose columns are the vectors Γ , and the Gram matrix of Γ is $B^T B$.

Theorem 13. *The Gram matrix corresponding to any basis of \mathbb{R}^n is symmetric and positive definite.*

Proof. Let A be the Gram matrix corresponding to some basis $\Omega = \omega_1, \dots, \omega_n$ of \mathbb{R}^n . Then A is symmetric because the inner product is commutative. To prove that A is positive definite, let $v \in \mathbb{R}^n$ be nonzero. Let B be the matrix whose columns are the vectors Ω . Then $A = B^T B$, so (considering v as a column vector) $v^T A v = (Bv)^T Bv = \langle Bv, Bv \rangle > 0$. Thus, A is positive definite. \square

We will call a matrix in $\text{GL}_n(\mathbb{R})$ *integral* if all its entries are integral. Recall that the group $\text{GL}_n(\mathbb{Z})$ is the group of integral matrices whose inverse is also integral. We now give a useful characterization of $\text{GL}_n(\mathbb{Z})$.

Lemma 3. *The group $\mathrm{GL}_n(\mathbb{Z})$ consists of the integral elements of $\mathrm{GL}_n(\mathbb{R})$ whose determinant is ± 1 .*

Proof. Suppose $A \in \mathrm{GL}_n(\mathbb{Z})$. Since $AA^{-1} = I$, taking determinants yields $(\det A)(\det A^{-1}) = 1$. Now both determinants are integers, so $\det A$ is invertible in \mathbb{Z} . Thus $\det A = \pm 1$. Conversely, suppose $A \in \mathrm{GL}_n(\mathbb{R})$ is integral and that $\det A = \pm 1$. To show that A^{-1} is integral, write the characteristic polynomial of A as

$$x^n + c_{n-1}x^{n-1} + \cdots + c_1x + (-1)^n(\det A).$$

By the Cayley–Hamilton theorem,

$$A^n + c_{n-1}A^{n-1} + \cdots + c_1A + (-1)^n(\det A)I = 0.$$

Now, multiplying by A^{-1} , we deduce that

$$(-1)^n(\det A)A^{-1} = -(A^{n-1} + c_{n-1}A^{n-2} + \cdots + c_1I).$$

Since the expression on the right is certainly an integral matrix, our assumption about the value of $\det A$ implies that A^{-1} is integral. \square

While the Gram matrix of a lattice L depends on the basis chosen, the determinant of the Gram matrix does not depend on the particular basis. To see this, let A be the Gram matrix of L with respect to the basis $\Omega = \omega_1, \dots, \omega_n$, and suppose that $\Gamma = \gamma_1, \dots, \gamma_n$ is another basis of \mathbb{R}^n that generates L . Let B be the change of basis matrix from Γ to Ω . Since each γ_i is a \mathbb{Z} -linear combination of the vectors $\omega_1, \dots, \omega_n$, the matrix B has integer entries. Similarly, the inverse of B also has integer entries. Thus, $B \in \mathrm{GL}_n(\mathbb{Z})$, and so $\det B = \pm 1$. Therefore, $\det B^T A B = \det A$. But $B^T A B$ is the Gram matrix of Γ . This shows that each possible Gram matrix of L has the same determinant, and we call this value the *discriminant* of L , which we will denote $\mathrm{disc} L$. The discriminant must be positive, since the Gram matrix is positive definite.

While a lattice does not determine a Gram matrix, we might ask whether a (candidate) Gram matrix uniquely determines a lattice. First, note that every symmetric positive definite matrix A is the Gram matrix of some lattice. This holds because A has a symmetric positive definite square root B , and then the columns of B form a basis for a lattice whose Gram matrix is $B^T B = B^2 = A$. Now, suppose we take a lattice L_1 with basis (the columns of) a matrix B . We want to find all other lattices (if any) that have the same Gram matrix as L_1 . Consider another lattice L_2 and let C be the change of basis matrix from a basis of L_2 to B . Then to say that the two lattices have equal Gram matrices is to say that $(CB)^T CB = B^T B$, which is equivalent to $C^T C = I$. That is, the inverse and transpose of C coincide, which means (by definition) that C is orthogonal. The conclusion, then, is that two lattices, generated by chosen bases, have the same Gram matrix if and only if the change of basis matrix from one basis to the other is orthogonal.

Thus, a Gram matrix determines multiple lattices. However, if two lattices have the same Gram matrix, the change of basis matrix B from one basis to another preserves inner products between the basis vectors, and hence between all vectors. Thus, in particular, B preserves the lengths of vectors and the angles between them, and the former property indicates that B is an isometry. We can then say that the two lattices are the same up to “orientation”.

Typically, we do not care about the orientation of a lattice, as the orientation is considered an artifact of the coordinates in \mathbb{R}^n . This motivates a different way to describe lattices. We consider that the lattice points are always just \mathbb{Z}^n , and that it is the inner product, specified by the Gram matrix, that varies. This perspective bundles in the notion that the orientation of a lattice is immaterial. Importantly, all the properties of lattices that we care about (such as the discriminant) can be determined from the Gram matrix.

Definition 7. A lattice is *integral* if its Gram matrix is integral. A lattice is *unimodular* if its discriminant is 1.

Note that the definition of an integral lattice makes sense; the condition is equivalent to asserting that the inner product of any two elements of the lattice is an integer.

If $\omega_1, \dots, \omega_n$ is a basis of \mathbb{R}^n , then the dual basis $\omega_1^*, \dots, \omega_n^*$ is the basis defined by the conditions $\langle \omega_i, \omega_i^* \rangle = 1$ and $\langle \omega_i, \omega_j^* \rangle = 0$ for $1 \leq i \neq j \leq n$. (It is easy to show that the vectors $\omega_1^*, \dots, \omega_n^*$ are linearly independent.)

Definition 8. Let L be a lattice in \mathbb{R}^n . The *dual lattice* to L , denoted L^* , is defined by

$$L^* = \{v \in \mathbb{R}^n : \forall \omega \in L \langle v, \omega \rangle \in \mathbb{Z}\}.$$

The following theorem shows that L^* is indeed a lattice.

Theorem 14. Let $\omega_1, \dots, \omega_n$ be a basis of \mathbb{R}^n , and let L be the lattice generated by this basis. Then L^* is the lattice generated by the dual basis of $\omega_1, \dots, \omega_n$.

Proof. Take a vector $v \in \mathbb{R}^n$. To have $v \in L^*$, it suffices to have $\langle v, \omega_i \rangle \in \mathbb{Z}$ for all $1 \leq i \leq n$. Write v as a linear combination of the dual basis vectors: $v = v_1 \omega_1^* + \dots + v_n \omega_n^*$. Now, using the definition of the dual basis, we see that for $1 \leq i \leq n$, we have $\langle v, \omega_i \rangle = v_i$. So, $v \in L^*$ if and only if v is an integer linear combination of the dual basis vectors. \square

From the definition, we see that the dual to a dual basis is the original basis. Therefore, the previous theorem shows that $L^{**} = L$. We next want to determine the Gram matrix of L^* in terms of that for L .

Theorem 15. Let A be the Gram matrix of a basis of \mathbb{R}^n . Then the dual basis has Gram matrix A^{-1} .

Proof. Take a basis $\omega_1, \dots, \omega_n$ of \mathbb{R}^n with Gram matrix A . Let B be the Gram matrix of the dual basis. We will show that $AB = I$. To do this, we compute the i, j entry of AB for all $1 \leq i, j \leq n$. Write

$$\omega_j^* = a_1 \omega_1 + \dots + a_n \omega_n \tag{*}$$

with $a_1, \dots, a_n \in \mathbb{R}$. Then, for any $1 \leq k \leq n$, we have $\langle \omega_k^*, \omega_j^* \rangle = a_k$. Thus,

$$(AB)_{i,j} = \sum_{k=1}^n \langle \omega_i, \omega_k \rangle \langle \omega_k^*, \omega_j^* \rangle = \sum_{k=1}^n a_k \langle \omega_i, \omega_k \rangle.$$

But, taking the inner product of (*) with ω_i yields

$$\sum_{k=1}^n a_k \langle \omega_i, \omega_k \rangle = \langle \omega_i, \omega_j^* \rangle,$$

which is 1 if $i = j$ but 0 if $i \neq j$. □

An immediate corollary is that the discriminant of the dual lattice L^* is the reciprocal of the discriminant of L . Now, a lattice is *self-dual* if $L = L^*$. We can characterize self dual lattices as follows.

Theorem 16. *A lattice is self-dual if and only if it is integral and unimodular (that is, if the Gram matrix of the lattice is in $\text{GL}_n(\mathbb{Z})$).*

Proof. Let L be a lattice. Suppose L is self-dual. Then it is immediate from Definition 8 that the inner product of any two lattice elements is an integer, so L is integral. Moreover, L and L^* have the same discriminant, but the discriminants are reciprocals of each other. Thus L is unimodular.

Conversely, suppose L is integral and unimodular. Since L is integral, we have $L \subseteq L^*$. If A is a Gram matrix of L , then $A \in \text{GL}_n(\mathbb{Z})$, so that $A^{-1} \in \text{GL}_n(\mathbb{Z})$. Thus, L^* is integral, so that $L^* \subseteq L^{**} = L$. □

4.2 Theta series

We now define theta series. Under certain conditions, these will turn out to be modular forms.

Let L be a lattice. For any real number k , let $N_k(L)$ be the number of vectors in L such that $\langle v, v \rangle = k$. (That there are finitely many such vectors will be proved below.) Then, for $z \in H$, we define the *theta series* of L to be

$$\Theta_L(z) = \sum_{k \geq 0} N_{2k}(L) q^k = \sum_{v \in L} q^{\langle v, v \rangle / 2},$$

where $q^k = e^{2k\pi iz}$. (In general, the k occurring in the sum can be arbitrary nonnegative real numbers, but in the case we will be interested in they will turn out to be integers.) In order to prove the convergence of the theta series, we will need the following lemma. This lemma, together with a similar statement involving the greatest eigenvalue, is known as Rayleigh's inequality.

Lemma 4. *Let A be a symmetric real matrix of dimension $n \times n$, and let λ be the smallest eigenvalue of A . Then, for every $v \in \mathbb{R}^n$,*

$$\lambda \langle v, v \rangle \leq v^T A v.$$

Proof. Since A is symmetric, it has an orthonormal basis of eigenvectors. Let $\omega_1, \dots, \omega_n$ be these eigenvectors, and let $\lambda_1, \dots, \lambda_n$ be the corresponding eigenvalues. Now write

$$v = a_1 \omega_1 + \dots + a_n \omega_n$$

with $a_1, \dots, a_n \in \mathbb{R}$. Then,

$$\begin{aligned} v^T A v &= v^T (a_1 \lambda_1 \omega_1 + \dots + a_n \lambda_n \omega_n) \\ &= a_1 \lambda_1 \langle v, \omega_1 \rangle + \dots + a_n \lambda_n \langle v, \omega_n \rangle \\ &= a_1^2 \lambda_1 + \dots + a_n^2 \lambda_n \\ &\geq \lambda (a_1^2 + \dots + a_n^2) \\ &= \lambda \langle v, v \rangle. \end{aligned}$$

(In the third and fifth lines we use the fact that the basis is orthonormal.) □

Theorem 17. *For every lattice L in \mathbb{R}^n , the number of vectors $v \in L$ such that $\langle v, v \rangle \leq k$ is $O(k^{n/2})$ (where k varies over the nonnegative real numbers). In particular, $N_k(L)$ is finite.*

Proof. Let B be a matrix whose columns are the vectors of a basis for L . Then, B , considered as a linear transformation, gives a bijection from \mathbb{Z}^n onto L , so that we are reduced to counting the number of $u \in \mathbb{Z}^n$ such that $\langle Bu, Bu \rangle \leq k$. Moreover, $A = B^T B$ is the Gram matrix for L . Let λ be the smallest eigenvalue of A . Since A is positive definite, $\lambda > 0$. Using the previous lemma, we obtain,

$$\langle Bu, Bu \rangle = (Bu)^T Bu = u^T A u \geq \lambda \langle u, u \rangle.$$

Thus, if we write $u = (u_1, \dots, u_n)$, we see that if $\lambda u_i^2 > k$ for some $1 \leq i \leq n$, then $\langle Bu, Bu \rangle > k$.

Hence, the number of u such that $\langle Bu, Bu \rangle \leq k$ is at most $(2\sqrt{k/\lambda} + 1)^n$. Since λ does not depend on k , this establishes the result. \square

Now, this is enough to prove that the theta series for \mathbb{Z}^n converges. Since \mathbb{Z}^n is integral, we need to take the sum only over k with $2k \in \mathbb{Z}$, and then it is easy to prove that the theta series converges absolutely for $z \in H$ (using, for instance, the ratio test). We can then generalize this result as follows.

Theorem 18. *The theta series for any lattice converges absolutely for $z \in H$.*

Proof. We use the notation in the previous theorem. Take $z \in H$, and let $r = q^\lambda$. Note that $|r| < 1$. Then,

$$\sum_{v \in L} |q|^{\langle v, v \rangle / 2} = \sum_{u \in \mathbb{Z}^n} |q|^{\langle Bu, Bu \rangle / 2} \leq \sum_{u \in \mathbb{Z}^n} |q|^{\lambda \langle u, u \rangle / 2} = \sum_{u \in \mathbb{Z}^n} |r|^{\langle u, u \rangle / 2},$$

which we already know converges. \square

In order to show that, under the right circumstances, theta series are modular forms, we need the following result. We give only a summary of the proof.

Theorem 19. *Suppose L is a lattice in \mathbb{R}^n and that z lies on the positive imaginary axis. Then,*

$$(\text{disc } L)^{1/2} (z/i)^{-n/2} \Theta_{L^*}(z) = \Theta_L(-1/z).$$

Proof sketch. We define a Schwartz function to be an infinitely differentiable function f from \mathbb{R}^n to \mathbb{C} that decays “quickly”. More precisely, for all real k , we must have that f and all partial derivatives of f are $o(\langle v, v \rangle^k)$ as the magnitude of v increases without bound. Then the Fourier transform of f , denoted \hat{f} , is given by

$$\hat{f}(u) = \int_{v \in \mathbb{R}^n} f(v) e^{2\pi i \langle v, u \rangle} dv$$

for all $v \in \mathbb{R}$. With this apparatus, one shows, for any lattice L ,

$$\sum_{v \in L} f(v) = (\text{disc } L)^{-1/2} \sum_{u \in L^*} \hat{f}(u).$$

(See [Elkies] for a proof of this.) Then, let t be a positive real number, and consider $f(v) = e^{-\pi\langle v,v\rangle/t}$. This is a Schwartz function. In order to compute the Fourier transform of f , we use the integral

$$\int_{\mathbb{R}} e^{-\pi x^2/t} e^{2\pi i x y} dx = t^{1/2} e^{-\pi t y^2}.$$

Write $v = (v_1, \dots, v_n)$ and $u = (u_1, \dots, u_n)$ in terms of their components under the standard basis. We find that

$$\begin{aligned} \hat{f}(u) &= \int_{v \in \mathbb{R}^n} e^{-\pi\langle v,v\rangle/t} e^{2\pi i\langle v,u\rangle} dv \\ &= \int_{v \in \mathbb{R}^n} \prod_{j=1}^n e^{-\pi v_j^2/t} e^{2\pi i v_j u_j} dv_j \\ &= \prod_{j=1}^n \int_{\mathbb{R}} e^{-\pi v_j^2/t} e^{2\pi i v_j u_j} dv_j \\ &= \prod_{j=1}^n t^{1/2} e^{-\pi t u_j^2} \\ &= t^{n/2} e^{-\pi t\langle u,u\rangle}. \end{aligned}$$

(The third line is obtained by rewriting the integral as nested one-dimensional integrals and moving constant factors out of integrations.) Applying our general result for Schwartz functions, we get

$$\sum_{v \in L} e^{-\pi\langle v,v\rangle/t} = (\text{disc } L)^{-1/2} t^{n/2} \sum_{u \in L^*} e^{-\pi t\langle u,u\rangle}.$$

Finally, by making the substitution $t = z/i$, we derive the desired result. □

Corollary 3. *Suppose the lattice L in \mathbb{R}^n is self-dual. Then, for all $z \in H$,*

$$(z/i)^{n/2} \Theta_L(z) = \Theta_L(-1/z).$$

Proof. A self-dual lattice is unimodular, so the previous theorem becomes the desired equation. This establishes the result when z is on the positive imaginary axis. Moreover, since the equations relates

analytic functions, the uniqueness of analytic continuations implies that the equation holds for all $z \in H$. \square

We can now proceed fairly quickly to the main result. Suppose the lattice L is self-dual. Then, L is integral, so the k occurring in the theta series are integers or half integers. Thus, the theta series are invariant under the map $z \mapsto z + 2$. Of course, to make the theta series modular forms for $\mathrm{SL}_2(\mathbb{Z})$, we need to have them invariant under $z \mapsto z + 1$, which we can achieve by imposing the further restriction that the lattice (or Gram matrix) is *even*. This means, by definition, that $\langle v, v \rangle$ is even for all $v \in L$, which is equivalent to asserting that the diagonal entries of the Gram matrix are even. Requiring L to be even forces the k in the theta series to be integers, so that the theta series are indeed invariant under $z \mapsto z + 1$.

Theorem 20. *Let L be an even self-dual lattice in \mathbb{R}^n . Then $8 \mid n$.*

Proof. For any $w \in H$, we apply Corollary 3 and the identity $\Theta_L(w) = \Theta_L(w + 1)$ to deduce that $\Theta_L(TSw) = (w/i)^{n/2}\Theta_L(w)$. Applying this three times (to $w = z$, TSz , and $(TS)^2z$), we get

$$\Theta_L(z) = \Theta_L((TS)^3z) = (((TS)^2z)/i)^{n/2}((TSz)/i)^{n/2}(z/i)^{n/2}\Theta_L(z),$$

the first equality holding because $(TS)^3 = -I$ acts trivially.

Now, let $z = i$ in the above. We have $TSi = i + 1$ and $(TS)^2i = (1 + i)/2$. Also, from the definition we see that $\Theta_L(i)$ is a positive real number, so we may cancel it from the equation. Thus, as $i + 1 = \sqrt{2}e^{\pi i/4}$,

$$1 = ((i + 1)/2)^{n/2}(i + 1)^{n/2} = (2^{-n/4}e^{\pi in/8})(2^{n/4}e^{\pi in/8}) = e^{\pi in/4},$$

which holds only when n is divisible by 8. \square

Theorem 21. *If L is an even, self-dual lattice in \mathbb{R}^n then Θ_L is a modular form of weight $n/2$.*

Proof. We already know that for Θ_L the matrix T satisfies the transformation rule for modular forms. The previous theorem implies that $i^{n/2} = 1$, so Corollary 3 simplifies to $\Theta_L(-1/z) = z^{n/2}\Theta_L(z)$,

which shows that S satisfies the transformation rule. Finally, Θ_L is holomorphic since it is given by a Fourier series with nonnegative powers. \square

4.3 Examples of even, self-dual lattices

In the last section we showed that theta series of even, self-dual lattices are modular forms, and along the way we saw that such lattices can only occur in spaces whose dimension is a multiple of 8. We can show that the converse is true; there is an even, self-dual lattice in \mathbb{R}^n for every n a multiple of 8. An example is the 8-dimensional lattice S_8 given by the Gram matrix

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

(To show that this matrix is the Gram matrix of a lattice, we need to show that it is positive definite, which we may do by verifying that all the minors of the matrix have positive determinant.)

From this example, we can construct even, self-dual lattices whose dimension is any multiple of 8 by forming a block diagonal matrix with copies of S_8 . Another example is the 24-dimensional lattice known as the Leech lattice. To define this lattice, first let B be the 11×11 matrix such that the i, j entry is 0 if $i = j$, and otherwise is 1 or -1 according as $i - j$ is or is not a square modulo 11. Then define a 12×12 matrix

$$A = \begin{pmatrix} 0 & E \\ -E^\top & B \end{pmatrix},$$

where E is the 1×11 matrix with every entry 1. More explicitly, A is the matrix

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 \\ -1 & 1 & 0 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 0 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & 0 & -1 & 1 & -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 & 0 & -1 & 1 & -1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 1 & -1 & 1 & 0 & -1 & 1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & -1 & 1 & 0 & -1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & 0 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & 0 & -1 & 1 \\ -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & 0 & -1 \\ -1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & 0 \end{pmatrix}.$$

And now the Leech lattice Λ_{24} is the lattice with Gram matrix

$$\begin{pmatrix} 4I & A - 2I \\ A^T - 2I & 4I \end{pmatrix},$$

where I is the 12×12 identity matrix. These are certainly not the only examples. While the lattice S_8 is in fact the only 8-dimensional even, self-dual lattice (up to isometry), there are 2 such lattices of dimension 16, and 24 of dimension 24. The number for higher multiples of 8 has not been determined, but it is known that there are at least 10^9 such lattices of dimension 32 (N. J. A. Sloane, Ed., The On-Line Encyclopedia of Integer Sequences, <http://www.research.att.com/~njas/sequences/A054909>).