MODULAR FORMS: THE VALENCE FORMULA AND SOME DIMENSION FORMULAE

Fan Zhou
fanzhou@college.harvard.edu

In this expository note we will discuss modular forms as symmetric differential forms on a Riemann surface[1]. We will assume the Riemann-Roch theorem and use it to prove the valence formula (though we also give a basic proof) and also some dimension formulae. The sources used are Serre's "A Course in Arithmetic" and Silverman's "Advanced Topics in the Arithmetic of Elliptic Curves", from whom I shall frequently appropriate diagrams.

Contents

1. The Modular Group

We will first explain what the modular group is and how it acts on the upper-half plane.

Let $\mathrm{SL}_2(\mathbb{R})$ act on $\mathbb{C} \cup \{\infty\}$ via

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z := \frac{az+b}{cz+d},$$

where we define the image to be ∞ when the denominator vanishes.

For $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2 \mathbb{R}$, it is easy to see

$$\operatorname{im} \frac{az+b}{cz+d} = \frac{\frac{az+b}{cz+d} - \frac{a\overline{z}+b}{c\overline{z}+d}}{2i} = \frac{acz\overline{z} + bc\overline{z} + adz + bd - acz\overline{z} - ad\overline{z} - bcz - bd}{2i|cz+d|^2} = \frac{(ad-bc)(z-\overline{z})}{2i|cz+d|^2},$$

so that using $ad - bc = 1$ we conclude

$$\operatorname{im}(gz) = \frac{\operatorname{im} z}{|cz+d|^2}.$$

This affords us that

$$\mathbb{H} := \{z \in \mathbb{C} : \operatorname{im} z > 0\}$$

[1]Here's something to tickle one's brain: we have RiemannIAN manifolds (with the Riemannian metric g thing), but just Riemann surfaces. Why don't we call them Riemannian surfaces? Or Riemann manifolds? Similarly, why are homomorphisms called that way, but holomorphic maps can't be holomorphisms? I cannot sleep.

is stable under the action of $\mathrm{SL}_2 \mathbb{R}$. Note also that $-\mathrm{Id} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ acts trivially, so it makes sense to consider

Definition. Let

$$G \coloneqq \mathrm{SL}_2 \mathbb{Z}/\{\pm \mathrm{Id}\} \coloneqq \mathrm{PSL}_2 \mathbb{Z}$$

be the *modular group.*

We may switch between using G and $\mathrm{PSL}_2 \mathbb{Z}$.

It turns out that this group is generated in a nice way:

Proposition. $\mathrm{PSL}_2 \mathbb{Z}$ is generated by

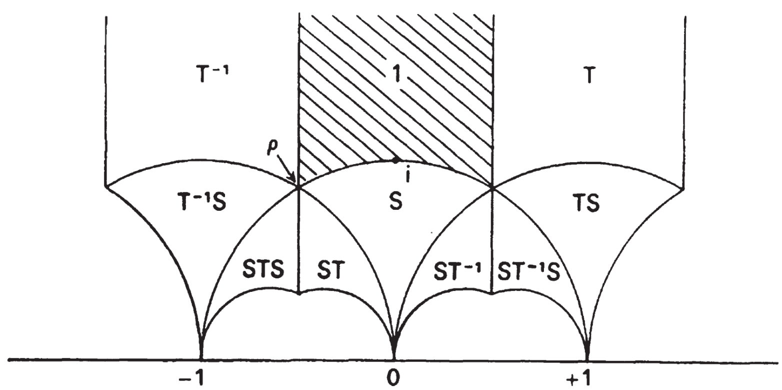
$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

$\square$

Note that S, T here act by

$$Tz = z + 1, \quad Sz = -\frac{1}{z}.$$

Instead of proving this theorem directly, we will consider the action of this group on the upper-half plane, following Serre. We will first show some nice facts about the shaded region below (diagram taken from Serre):



We will denote this shaded region, a so-called "fundamental domain", by

$$D \coloneqq \left\{ z \in \mathbb{H} : -\frac{1}{2} \leq \operatorname{re} z \leq \frac{1}{2} \right\} \cap \{ z \in \mathbb{H} : |z| \geq 1 \}.$$

Note that this contains its boundary.

Theorem.

- $\forall\, z \in \mathbb{H},\ \exists\, g \in \mathrm{PSL}_2\,\mathbb{Z} : gz \in D$.
- If $z \neq w$ are the same mod $\mathrm{PSL}_2\,\mathbb{Z}$, then either
  * $\operatorname{re} z = \pm\frac{1}{2}$ and $w = z \mp 1$, or
  * $|z| = 1$ and $w = -\frac{1}{z}$.
- For each $z \in D$, let $\mathrm{Stab}(z) := \{g \in G : gz = z\}$ denote the stabilizer. Then, for $z \in D$,
$$\mathrm{Stab}(z) = \{\mathrm{Id}\} \qquad \forall\, z \neq i, \omega_3, -\overline{\omega}_3,$$
where $\omega_3 = e^{2\pi i/3}$. The exceptions are
$$\begin{aligned}\mathrm{Stab}(i) &= \{\mathrm{Id}, S\},\\ \mathrm{Stab}(\omega_3) &= \{\mathrm{Id}, ST, (ST)^2\},\\ \mathrm{Stab}(-\overline{\omega}_3) &= \{\mathrm{Id}, TS, (TS)^2\}.\end{aligned}$$

□

Proof of Theorem and Proposition. • First we show that any member of $\mathbb{H}$ can be brought to D via some g. Let
$$G' = \langle S, T\rangle$$
be the subgroup of $G = \mathrm{PSL}_2\,\mathbb{Z}$ generated by S, T. We will show that in fact it can be done by $g \in G'$. Recalling
$$\operatorname{im}(gz) = \frac{\operatorname{im} z}{|cz+d|^2},$$
and noting that $c, d \in \mathbb{Z}$ implies there are only finitely many pairs (c, d) such that $0 < |cz+d|^2 \leq C$, we can take a $g \in G'$ such that $\operatorname{im}(gz)$ is maximized and finite. Then apply T enough times so that
$$-\frac{1}{2} \leq \operatorname{re} T^n gz \leq \frac{1}{2}.$$
If this had $|T^n gz| < 1$, then $-\frac{1}{T^n gz}$ would have strictly larger imaginary part than that of $T^n gz$, i.e. that of gz, contradicting maximality. Hence we conclude $|T^n gz| \geq 1$, so that
$$T^n gz \in D,$$
which gives the desired construction.

• Next we show the stabilizing and boundary gluing properties. Suppose w, z are equivalent mod $\mathrm{PSL}_2\,\mathbb{Z}$; WLOG let $\operatorname{im} w \geq \operatorname{im} z$, and let $w = gz$, so that $\operatorname{im} w = \operatorname{im} gz = \frac{\operatorname{im} z}{|cz+d|^2} \geq \operatorname{im} z$, or
$$|cz + d| \leq 1.$$
Since $|z| \geq 1$, this is clearly impossible for $|c| \geq 2$. It suffices to consider $c = -1, 0, 1$.

* If $c = 0$, then $ad - bc = 1$ requires $a, d = \pm 1$, so that g acts by translation by $\pm b$. As $-\frac{1}{2} \leq \operatorname{re}(gz), \operatorname{re}(z) \leq \frac{1}{2}$, this means either $b = 0$ (which gives $g = 1$) or $b = \pm 1$, in which case $\operatorname{re}(z), \operatorname{re}(gz) = \pm\frac{1}{2}$ (one is plus and the other is minus), which gives the first case in the theorem.
* If $c = 1$, then the condition becomes
$$|z + d| \leq 1$$
which forces $d = 0$ unless $z = \omega_3$ or $-\overline{\omega}_3$, in which case $d = 0, 1$ or $d = 0, -1$.

- * If $d=0$, the condition says $|z| \le 1$; but $z \in D$, so $|z|=1$. $ad-bc=1$ then requires $b=-1$, so that
$$gz = \frac{az+b}{z} = a - \frac{1}{z}.$$
In order for this to still be in D (note $-\frac{1}{z}$ is also on the boundary of the circle), a must be zero unless $z = \omega_3, -\overline{\omega}_3$, which we address later. For $a=0$, we then have $g = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, which gives the second case of the theorem.
- * If $z=\omega_3$ and $d=0, a=-1$, then[2] $g = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix} = (ST)^2$, which sends $gz=z$. In this case we find the stabilizer of ω_3 is at least as advertised.
- * If $z=-\overline{\omega}_3$ and $d=0, a=1$, then $g = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} = TS$, which sends $gz=z$. In this case we find the stabilizer of $-\overline{\omega}_3$ is at least as advertised.
- * If $z=\omega_3$ and $d=1$, $ad-bc=1$ gives $a-b=1$ and $g\omega_3 = \frac{a\omega_3+(a-1)}{\omega_3+1} = a - \frac{1}{\omega_3+1} = a+\omega_3$, so either $a=0$ in which case $g = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} = ST$ is again in the advertised stabilizer or $a=1$ in which case $g\omega_3 = -\overline{\omega}_3$, in which case $-\overline{\omega}_3 = w = -\frac{1}{z}$.
- * If $z=-\overline{\omega}_3$ and $d=-1$, $ad-bc=1$ gives $-a-b=1$ and $g(-\overline{\omega}_3) = a - \frac{1}{-\overline{\omega}_3-1} = a-\overline{\omega}_3$, so either $a=0$ in which case $g = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} = (TS)^2$ is again in the advertise stabilizer or $a=-1$ in which case $g(-\overline{\omega}_3) = \omega_3$, in which case $w = -\frac{1}{z}$ again.

* If $c=-1$, then switch the sign on all of a,b,c,d (recall $\mathrm{PSL}_2\,\mathbb{Z}$ is $\mathrm{SL}_2\,\mathbb{Z}$ modded out by switching signs) to return to the first case.

By reviewing the above cases for what g's fix each z, we obtain the stabilizers advertised.

Lastly, we show that $\langle S, T\rangle = G'$ is actually all of G. For any $g \in G$, consider $g(2i)$ (where $2i \in D$); by the first claim we proved there is some $g' \in G'$ such that $g'(g(2i)) \in D$ as well. Then $2i$ and $g'g(2i)$ are congruent mod G, one of which is in the interior of D; by the above this gives $g'g = \mathrm{Id}$, so $g \in G'$, so $G' = G$, and we conclude $\mathrm{PSL}_2\,\mathbb{Z}$ is generated by S and T. ■

Note that this means $\mathrm{Stab}(z)$ for $z \in \mathbb{H}$ is finite and cyclic.

It makes sense to speak of $\mathbb{H}/\mathrm{PSL}_2\,\mathbb{Z}$, which we will sometimes call $\mathbb{H}/G$ because it is shorter. The intuitive picture we get for this quotient from the above geometric insight is that the two vertical sides of the fundamental domain D are identified via $\mathrm{PSL}_2\,\mathbb{Z}$, as are the arcs going from ω_3 and $-\overline{\omega}_3$ to i; folding this together, we then get a "hot pocket" of sorts, or a topological sphere minus a point.

2. The Modular Curve

The end goal is to endow something like $\mathbb{H}/G$ with the structure of a Riemann surface and do complex analysis.

Definition. Let
$$\widehat{\mathbb{H}} = \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q}) = \mathbb{H} \cup \mathbb{Q} \cup \{\infty\}$$
be the *extended upper-half plane*, where $\mathbb{Q}$ is thought of as laying along the real axis in the complex plane of which $\mathbb{H}$ is the upper half. The points of $\mathbb{P}^1(\mathbb{Q})$ are called *cusps* of $\widehat{\mathbb{H}}$.

[2] I will skip the matrix multiplication verification as it is just algebra.

We had earlier defined the action of $\mathrm{PSL}_2\,\mathbb{Z}$ on $\mathbb{H}$; we extend this to $\widehat{\mathbb{H}}$ by letting the group act on $\mathbb{P}^1(\mathbb{Q})$ via

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix},$$

where $(x : y) = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{P}^1(\mathbb{Q})$ are homogeneous coordinates. Under this action of $\mathrm{PSL}_2\,\mathbb{Z}$ on $\widehat{\mathbb{H}}$, we may consider

$$X \coloneqq \widehat{\mathbb{H}}/\mathrm{PSL}_2\,\mathbb{Z}.$$

We now extend the geometric insight from last section to this object.

Proposition. Letting

$$X = \widehat{\mathbb{H}}/\mathrm{PSL}_2\,\mathbb{Z},$$

we have

$$X \setminus (\mathbb{H}/\mathrm{PSL}_2\,\mathbb{Z}) = \{\infty\}$$

and

$$\mathrm{Stab}(\infty) = \left\{\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}\right\} = \langle T\rangle,$$

the subgroup generated by translation.

□

Proof. To show that $X \setminus (\mathbb{H}/G) = \{\infty\}$, it suffices to see that every point in $\mathbb{P}^1(\mathbb{Q})$ can be brought to ∞ under the action of G. To this end let $\begin{pmatrix} x \\ y \end{pmatrix}$ be any such point; being homogeneous coordinates, we may as well assume $x, y \in \mathbb{Z}$ with $\gcd(x, y) = 1$. Then by Bezout there are $a, b \in \mathbb{Z}$ such that $ax + by = 1$; then

$$\begin{pmatrix} a & b \\ -y & x \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

exhibits the element of G bringing any $(x : y)$ to $(1 : 0) = \infty$.

For the stabilizer fact, observe

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

if and only if $c = 0$. Since the determinant should be 1, we get the matrix is of form $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$ as claimed. ■

We noted earlier that $\mathbb{H}/G$ was topologically a sphere minus a point; from our above conclusions, we may intuit that $X = \widehat{\mathbb{H}}/G$ is then a topological sphere and therefore has genus 0. To make this precise we endow $\widehat{\mathbb{H}}$ with a topological structure.

Definition. Let $\widehat{\mathbb{H}}$ be endowed with the topology whose open sets are generated by the following: usual open neighborhoods

$$\mathrm{Nbhd}(z) \qquad \text{for each } z \in \mathbb{H};$$

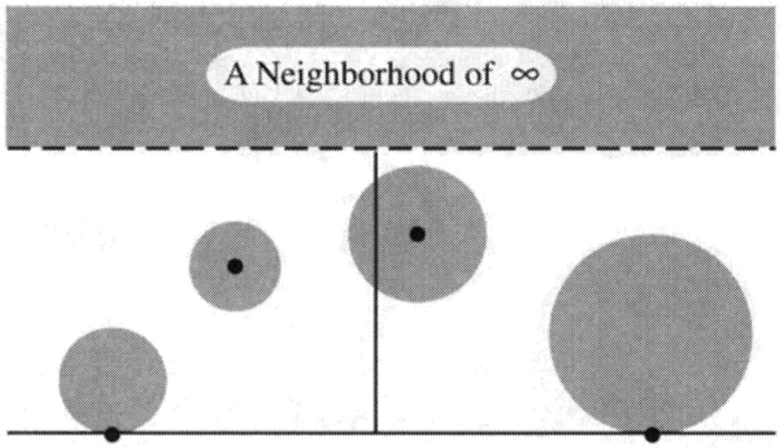
sets around ∞

$$\{z \in \mathbb{H} : \operatorname{im} z > B\} \cup \{\infty\} \qquad \text{for each } B > 0;$$

and sets around $\mathbb{Q}$

$$\mathrm{Int}(\text{circle in } \mathbb{H} \text{ tangent to the real axis at } r) \cup \{r\} \qquad \text{for each } r \in \mathbb{Q}.$$

Here is a picture from Silverman for intuition:



We saw earlier that there is always some fractional linear transformation g sending $g\infty = r$; one can moreover check that this g sends the neighborhoods $\{z \in \mathbb{H} : \operatorname{im} z > B\} \cup \{\infty\}$ to the circles $\operatorname{Int}(\text{circle in } \mathbb{H} \text{ tangent to the real axis at } r) \cup \{r\}$. It is also clear from the above definitions that $\widehat{\mathbb{H}}$ is a Hausdorff topological space.

$\widehat{\mathbb{H}}$ also enjoys the following property:

Proposition. Letting[4]

$$I(z_1, z_2) := \{g \in \mathrm{PSL}_2 \mathbb{Z} : gz_1 = z_2\}$$

for $z_1, z_2 \in \widehat{\mathbb{H}}$, and letting

$$I(U_1, U_2) := \{g \in \mathrm{PSL}_2 \mathbb{Z} : gU_1 \cap U_2 \neq \emptyset\}$$

for $U_1, U_2 \subseteq \widehat{\mathbb{H}}$, we have that for any $z_1, z_2 \in \widehat{\mathbb{H}}$, there exist neighborhoods $\mathrm{Nbhd}(z_1)$ and $\mathrm{Nbhd}(z_2)$ such that

$$I(\mathrm{Nbhd}(z_1), \mathrm{Nbhd}(z_2)) = I(z_1, z_2).$$

□

The proof of this first notes that it suffices to show this for $z_1, z_2 \in D \cup \{\infty\}$, and then breaks into casework on whether or not a point is infinity. We will skip this proof.

Given this topological structure for $\widehat{\mathbb{H}}$, we now have a topological structure for $X = \widehat{\mathbb{H}}/G$. Recall that the quotient topology is such that $U \subseteq X$ is open if and only if $\pi^{-1}(U)$ is open; alternatively, it is the weakest topology for which the projection map π is continuous. Note that π is also an open map since if $U \subseteq \widehat{\mathbb{H}}$ is open then so is $\pi^{-1}(\pi(U)) = \bigcup_{g \in G} gU$, which implies by definition that $\pi(U)$ is open. We move towards making X a Riemann surface.

First we claim

Proposition. X under this quotient topology is a compact Hausdorff space.

□

Proof. To see that X is compact, let $\{U_\alpha\}$ be any open cover of X; then $\{\pi^{-1}U_\alpha\}$ is an open cover of $\widehat{\mathbb{H}}$. Since X is covered, there must be some i for which $\infty \in \pi^{-1}U_i$; by the topology of $\widehat{\mathbb{H}}$ there is the $B > 0$ such that

$$\{z \in \mathbb{H} : \operatorname{im} z > B\} \cup \{\infty\} \subseteq \pi^{-1}U_i.$$

Therefore the set $D \setminus \pi^{-1}U_i$, being closed and bounded, is compact, and therefore $\{\pi^{-1}U_\alpha\}$ which covers (using the $X \subseteq \cup U$ definition; certainly we can intersect the covering sets if we want equality)

$D \setminus \pi^{-1}U_i$ must admit a finite subcover

$$D \setminus \pi^{-1}U_i \subseteq \bigcup_{j \text{ fnt}} \pi^{-1}U_j.$$

Then

$$X \subseteq U_i \cup \bigcup_{j \text{ fnt}} U_j.$$

To see that X is Hausdorff, for $p_1, p_2 \in X$ distinct points let z_1, z_2 be lifts to $\widehat{\mathbb{H}}$. Since $gz_1 \neq z_2$ for any $g \in G$, we know $I(z_1, z_2) = \emptyset$, which from our previous proposition means there is $U_1, U_2 \subseteq \widehat{\mathbb{H}}$ such that $I(U_1, U_2) = \emptyset$. Then $\pi(U_1), \pi(U_2)$ are disjoint open sets around p_1, p_2 in X. ■

Next we give X the structure of a Riemann surface.

Theorem. $X = \widehat{\mathbb{H}}/\mathrm{PSL}_2\,\mathbb{Z}$ has the structure of a compact connected Riemann surface of genus 0 via the following:

For $p \in X$, let z_p be a lift along $\pi\colon \widehat{\mathbb{H}} \twoheadrightarrow X$, i.e. such that $\pi(z_p) = p$. Let $U_{z_p} \subseteq \widehat{\mathbb{H}}$ be a neighborhood of z_p with

$$I(U_{z_p}, U_{z_p}) = \mathrm{Stab}(z_p);$$

this is possible from an earlier proposition by taking $z_1 = z_2 = z_p$ and $U_{z_p} = U_1 \cap U_2$ the intersection of the sets we were guaranteed there.

Then

$$U_p \coloneqq U_{z_p}/\mathrm{Stab}(z_p) \subseteq X$$

is a neighborhood of p, and $\{U_{z_p}/\mathrm{Stab}(z_p)\}_p$ gives an open cover of X. This open cover forms an atlas for the complex manifold X with the following charts:

For $p \neq \infty$, let[6]

$$\begin{aligned} \psi_{z_p}\colon \mathbb{H} &\longrightarrow B_1(0) \\ z &\longmapsto \frac{z - z_p}{z - \overline{z}_p} \end{aligned}$$

which is a biholomorphic map; then the chart for U_p is

$$\begin{aligned} \phi_p\colon U_p &\longrightarrow \mathbb{C} \\ p &\longmapsto \psi_{z_p}(\pi^{-1}(p))^{|\mathrm{Stab}\, z_p|}. \end{aligned}$$

Part of the claim is that this is well-defined.

For $p = \infty$, we may take $z_p = \infty$ with $\mathrm{Stab}(z_p) = \langle T \rangle$. Then the chart for U_∞ is

$$\begin{aligned} \phi_\infty\colon U_\infty &\longrightarrow \mathbb{C} \\ p &\longmapsto \begin{cases} e^{2\pi i \pi^{-1}(p)} & p \neq \infty \\ 0 & p = \infty \end{cases}. \end{aligned}$$

Part of the claim is that this is well-defined.

Note well that these charts all send p to zero:

$$\phi_p(p) = 0.$$

□

Here are some pictures[7] to illustrate the charts:

[7]Mostly because I can't figure out how to put tikzcd into the environment above.

$$\begin{tikzcd} U_{z_p} \arrow[r, two heads, "\pi"] \arrow[d, "\psi_{z_p}"'] & U_p = U_{z_p}/\operatorname{Stab} z_p \arrow[d, "\phi_p"] \\ \mathbb{C} \arrow[r, "\square^{|\operatorname{Stab} z_p|}"] & \mathbb{C} \end{tikzcd}$$

$$\begin{tikzcd} U_{z_\infty} \arrow[r, two heads, "\pi"] \arrow[dr, "\psi_\infty"'] & U_\infty = U_{z_\infty}/\langle T\rangle \arrow[d, "\phi_\infty"] \\ & \mathbb{C} \end{tikzcd}$$

where $\psi_\infty(z) := e^{2\pi i z}$.

Before proving the theorem, let us prove a quick lemma:

Lemma. For given $a \in \mathbb{H}$, let $F\colon \mathbb{H} \longrightarrow \mathbb{H}$ be such that $F(a) = a$, and let ψ_a be a function on $\mathbb{H}$

$$\psi_a(z) = \frac{z-a}{z-\overline{a}}.$$

If

$$F^{\circ m} = \mathrm{Id}$$

is the smallest such m, then there exists ω_m a primitive m-th root of unity such that

$$\psi_a \circ F = \omega_m \psi_a$$

on $\mathbb{H}$.

□

Proof of Lemma. Recall that ψ_a is a biholomorphism $\psi_a\colon \mathbb{H} \xrightarrow{\sim} B_1(0)$ with $\psi_a(a) = 0$. Note that $F^{\circ m} = \mathrm{Id}$ means that F is invertible and therefore biholomorphic with inverse $F^{\circ(m-1)}$. Then the map

$$\psi_a \circ F \circ \psi_a^{-1}\colon B_1(0) \xrightarrow{\sim} B_1(0)$$

is a biholomorphism of the unit disk to itself sending zero to zero. We saw in M213aps5 the fifth pset that, since

Fact.

$$\operatorname{Aut} B_1(0) = \left\{ e^{i\theta}\psi_\alpha(z) = e^{i\theta}\frac{\alpha - z}{1-\overline{\alpha}z} : |\alpha| < 1 \right\}.$$

□

we must have that

$$\psi_a \circ F \circ \psi_a^{-1} = \times e^{i\theta}$$

for some θ; moreover since $(\psi_a \circ F \circ \psi_a^{-1})^{\circ m} = \psi_a \circ F^{\circ m} \circ \psi_a^{-1} = \mathrm{Id}$ is the minimally chosen such m, we conclude that $e^{i\theta}$ must be a primitive m-th root of unity, as claimed. ■

We now proceed with the proof of the theorem.

Proof of Theorem. We already know X is compact Hausdorff, and we also know it is connected since it is the continuous image of a connected space, $\pi\colon \widehat{\mathbb{H}} \to X$. We argued earlier that it has genus 0 (under the quotient action of G, all the neighborhoods around $\mathbb{Q}$ and ∞ can be sent to each other; this gives the open sets around the north pole of the topological sphere).

It remains to check the complex structure. We first show that the ϕ's so defined are homeomorphisms, and we then check that the transition maps are holomorphic.

We begin by observing that most of the time $z_p \neq i, \omega_3, \infty$, and we saw in section 1 that in these cases $\mathrm{Stab}(z_p) = \{\mathrm{Id}\}$, so that immediately $\phi_p = \psi_{z_p}$ is a well-defined homeomorphism. (Perhaps this is easier to see with the diagram above.)

In the cases $z_p = i, \omega_3$, we saw in section 1 that $\operatorname{Stab} z_p$ is cyclic of orders 2 and 3 respectively. Then any non-identity element of $\operatorname{Stab} z_p$ is a generator; let this element be g. Then by the lemma,

$$\psi_{z_p}(gz) = \omega_{\operatorname{ord} g}\psi_{z_p}(z),$$

where $\operatorname{ord} g = |\operatorname{Stab} z_p|$. But since $\omega_{\operatorname{ord} g}^{|\operatorname{Stab} z_p|} = 1$, this then immediately means via the first commutative diagram above that

$$\phi_p \colon \square^{|\operatorname{Stab} z_p|} \circ \psi_{z_p} \circ \pi^{-1}$$

is independent of lifts. Hence ϕ_p is well-defined.

To check it is a homeomorphism, first note that in the diagram, $\pi, \psi_{z_p}, \square^{|\operatorname{Stab} z_p|}$ are all continuous and open maps (ψ_{z_p} is moreover homeomorphic), so we immediately have ϕ_p is locally homeomorphic. It hence suffices to check that ϕ_p is injective to show it is a homeomorphic. To this end, let $p_1, p_2 \in U_p$, and note (again letting g be a generator of cyclic $\operatorname{Stab} z_p$)

$$\begin{aligned}
\phi_p(p_1) = \phi_p(p_2) &\iff \psi_{z_p}(\pi^{-1}p_1)^{|\operatorname{Stab} z_p|} = \psi_{z_p}(\pi^{-1}p_2)^{|\operatorname{Stab} z_p|} \\
&\iff \psi_{z_p}(\pi^{-1}p_1) = \omega_{|\operatorname{Stab} z_p|}^k \psi_{z_p}(\pi^{-1}p_2) \quad \text{for some } k \\
&\iff \psi_{z_p}(\pi^{-1}p_1) = \psi_{z_p}(g^k\pi^{-1}p_2) \quad \text{for some } k \\
&\iff \pi^{-1}p_1 = g^k\pi^{-1}p_2 \quad \text{for some } k \\
&\iff p_1 = p_2
\end{aligned}$$

Hence ϕ_p is a well-defined homeomorphism.

In the case $p = \infty$, recall $\operatorname{Stab} \infty = \langle T \rangle$ is also cyclic where T^k is translation by k. Then

$$\phi_\infty \coloneqq \psi_\infty \circ \pi^{-1}$$

is also well-defined since $\psi_\infty(z) = e^{2\pi i z}$ and shifting z by $\mathbb{Z}$ does not change the answer. Since this translation is also precisely the kernel of $z \mapsto e^{2\pi i z}$, we conclude that ϕ_∞ is also injective, and by the same argument as above (each leg being continuous and open) we conclude ϕ_∞ is also a well-defined homeomorphism.

Lastly it remains to check holomorphicity of the transition maps.

First let $p, q \neq \infty$. Note that on the intersection,

$$\phi_p \circ \phi_q^{-1} = \square^{|\operatorname{Stab} z_p|} \circ \psi_{z_p} \circ \pi^{-1} \circ \pi \circ \psi_{z_q}^{-1} \circ \square^{\frac{1}{|\operatorname{Stab} z_q|}} = \square^{|\operatorname{Stab} z_p|} \circ \psi_{z_p} \circ \psi_{z_q}^{-1} \circ \square^{\frac{1}{|\operatorname{Stab} z_q|}},$$

where everything is holomorphic except $\square^{\frac{1}{|\operatorname{Stab} z_q|}}$. But this is fine, since $\psi_a(F(z)) = \omega_m\psi_a(z)$ also means $\psi_a^{-1}(\omega_m z) = F(\psi_a^{-1}(z))$ (plug in $z = \psi_a^{-1}(z)$ to see this), so that

$$\square^{|\operatorname{Stab} z_p|} \circ \psi_{z_p} \circ \pi^{-1} \circ \pi \circ \psi_{z_q}^{-1} \circ (\omega_{|\operatorname{Stab} z_q|}z) = \square^{|\operatorname{Stab} z_p|} \circ \psi_{z_p} \circ \pi^{-1} \circ \pi \circ g \circ \psi_{z_q}^{-1} \circ z = \square^{|\operatorname{Stab} z_p|} \circ \psi_{z_p} \circ \pi^{-1} \circ \pi \circ \psi_{z_q}^{-1} \circ z,$$

where the g is a generator for the cyclic $\operatorname{Stab} z_q$ and is absorbed by π. It follows that

$$\psi_{z_p}(\psi_{z_q}^{-1}(\omega_{|\operatorname{Stab} z_q|}z))^{|\operatorname{Stab} z_p|} = \psi_{z_p}(\psi_{z_q}^{-1}(z))^{|\operatorname{Stab} z_p|}$$

and therefore that

$$\psi_{z_p}(\psi_{z_q}^{-1}(z))^{|\operatorname{Stab} z_p|}$$

is a power series in $z^{|\operatorname{Stab} z_q|}$, meaning the transition map

$$\phi_p \circ \phi_q^{-1}(z) = \psi_{z_p}(\psi_{z_q}^{-1}(z^{1/|\operatorname{Stab} z_q|}))^{|\operatorname{Stab} z_p|}$$

is holomorphic.

Similarly, for $p \neq \infty$, one may find the transition map

$$\phi_\infty \circ \phi_p^{-1}(z) = e^{2\pi i \psi_{z_p}^{-1}(z^{1/|\operatorname{Stab} z_p|})}$$

is holomorphic.

In the case of

$$\phi_p \circ \phi_\infty^{-1},$$

first note that

$$\psi_{z_p}(z+1)^{|\operatorname{Stab} z_p|} = \phi_p(\pi(Tz)) = \phi_p(\pi(z)) = \psi_{z_p}(z)^{|\operatorname{Stab} z_p|},$$

which means $\psi_{z_p}(z)^{|\operatorname{Stab} z_p|}$ is a holomorphic function in the coordinate $q = e^{2\pi i z}$. Therefore the transition map

$$\phi_p \circ \phi_\infty^{-1}(z) = \psi_{z_p}\left(\frac{1}{2\pi i}\log z\right)^{|\operatorname{Stab} z_p|}$$

is holomorphic[8].

This completes the proof that the charts above grant X a complex structure. ■

Now that we have our Riemann surface X, we will proceed to study so-called modular forms on it.

3. Modular Functions from a Classical Approach

In the below $\mathscr{C}^+$ will denote holomorphic and $\mathscr{C}^{+/+}$ will denote meromorphic.

Definition. $f \in \mathscr{C}^{+/+}(\mathbb{H})$ is said to be *weakly modular of weight* 2k if, for any $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in$ $\mathrm{PSL}_2\,\mathbb{Z}$,

$$f(gz) = (cz+d)^{2k} f(z).$$

Noting that $\frac{\mathrm{d}}{\mathrm{d}z} gz = \frac{a(cz+d)-(az+b)c}{(cz+d)^2} = \frac{1}{(cz+d)^2}$, the above requirement can be written

$$\frac{f(gz)}{f(z)} = \left(\frac{\mathrm{d}(gz)}{\mathrm{d}z}\right)^{-k},$$

which is formally

$$f(gz)\,\mathrm{d}(gz)^k = f(z)\,\mathrm{d}z^k.$$

We will make this precise later.

Since $\mathrm{PSL}_2\,\mathbb{Z}$ is generated by S and T, to check the invariance of the above tensor form it suffices to check invariance on S, T. This is precisely saying

Proposition. For $f \in \mathscr{C}^{+/+}(\mathbb{H})$, f is a weakly modular function of weight $2k$ if and only if

$$f(z+1) = f(z),$$
$$f\left(-\frac{1}{z}\right) = z^{2k} f(z).$$

□

If a function satisfies $f(z+1) = f(z)$, then, using the change of variables

$$q = e^{2\pi i z},$$

[8]Here note that $z \in U_{z_p} \cap U_\infty$ so z is bound by U_{z_p} to not tend towards infinity.

or

$$z = \frac{\log q}{2\pi i},$$

we see that $f(z) = f\left(\frac{\log q}{2\pi i}\right)$ is well-defined despite the multi-valuedness of the logarithm due to the periodicity of f. Hence we can write f as

$$f(z) = f\left(\frac{\log q}{2\pi i}\right) = \widehat{f}(q),$$

where $q \neq 0$. Since f is meromorphic on $\mathbb{H}$, $\widehat{f}$ is meromorphic on the deleted disk $0 < |q| < 1$ (here we note $|q| = e^{-2\pi \operatorname{im} z}$, but $\operatorname{im} z > 0$, so $|q| < 1$).

Definition. When $\widehat{f}$ extends to a meromorphic (resp. holomorphic) function on $|q| < 1$, we say f is *meromorphic (resp. holomorphic) at infinity.*

A *modular function* is a weakly modular function which is meromorphic at infinity.

A *modular form* is a modular function which is holomorphic everywhere (including infinity).

A *cusp form* is a modular form which is zero at infinity.

f being meromorphic at infinity amounts to admitting a Laurent expansion

$$\widehat{f}(q) = \sum_{n=-N}^{\infty} a_n q^n;$$

being holomorphic at infinity means $N = 0$. In that case, we set

$$f(\infty) \coloneqq \widehat{f}(0).$$

A modular form of weight $2k$ is then

$$f(z) = \widehat{f}(q) = \sum_{n=0}^{\infty} a_n q^n = \sum_{n=0}^{\infty} a_n e^{2\pi i n z}$$

converging on $|q| < 1$ which moreover satisfies

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

Note that, if f, g are modular forms of weights $2k, 2l$, then their product fg is still holomorphic everywhere, satisfies $f(z+1)g(z+1) = f(z)g(z)$, and $f(-1/z)g(-1/z) = z^{2k} f(z) z^{2l} g(z) = z^{2(k+l)} f(z) g(z)$, so fg is a modular form of weight $2(k+l)$.

Similarly, if f, g are modular forms of the same weight, then so is their sum.

For examples of these things, skip to section 6.1.

4. Modular Functions as Forms on the Riemann Surface

We explore another way to think of things. We start with general considerations, for any Riemann surface X.

Definition. Let
$$\Omega^0(X) \coloneqq \mathbb{C}(X)$$
be the field of meromorphic functions on X, and let
$$\Omega^1(X)$$
be the space of meromorphic 1-forms on X. Then let
$$\Omega^k(X) \coloneqq (\Omega^1(X))^{\otimes k} = \Omega^1(X) \otimes_{\mathbb{C}(X)} \cdots \otimes_{\mathbb{C}(X)} \Omega^1(X).$$
be the space of symmetric meromorphic k-forms.

For such a form $\omega \in \Omega^k(X)$, we may choose a local coordinate z in some neighborhood of $p \in X$ and write
$$\omega = f \,\mathrm{d}z^k$$
for some meromorphic $f \in \Omega^0(X)$. Define
$$\operatorname{ord}_p \omega \coloneqq \operatorname{ord}_p f,$$
where $\operatorname{ord}_p f$ is defined as the order of vanishing of the function f. Recall that this number is independent of choice of chart. Also define a formal sum
$$\operatorname{div} \omega \coloneqq \sum_{p \in X} \operatorname{ord}_p(\omega) p,$$
with
$$\deg(\operatorname{div} \omega) \coloneqq \sum_{p \in X} \operatorname{ord}_p \omega.$$

Note that ω is moreover holomorphic precisely when $\operatorname{ord}_p \omega \geq 0$ at all points p. Recall for functions f on $\mathbb{C}$
$$\operatorname{ord}_p f$$
is the integer n such that $\frac{f}{(z-p)^n}$ is holomorphic and nonzero at p.

We may now return to our ventures in modular forms. We claim that a modular function on $\mathbb{H}$ of weight $2k$ gives rise to a symmetric meromorphic k-form $\omega_f \in \Omega^k(X)$, where $X = \widehat{\mathbb{H}}/\mathrm{PSL}_2\,\mathbb{Z}$ is the Riemann surface we were discussing in section 2. First, letting $\mathrm{d}z$ be the usual 1-form on $\mathbb{H}$, observe
$$\mathrm{d}(gz) = \mathrm{d}\frac{az+b}{cz+d} = \frac{ad-bc}{(cz+d)^2}\,\mathrm{d}z = (cz+d)^{-2}\,\mathrm{d}z,$$
so that in general
$$\mathrm{d}(gz)^k = (cz+d)^{-2k}\,\mathrm{d}z^k.$$
If f is moreover a modular function of weight $2k$, then
$$f(gz)\,\mathrm{d}(gz)^k = (cz+d)^{2k} f(z)(cz+d)^{-2k}\,\mathrm{d}z^k = f(z)\,\mathrm{d}z^k$$
is a k-form invariant under the action of $\mathrm{PSL}_2\,\mathbb{Z}$.

Note that $\mathbb{H}$ automatically has the structure of a Riemann surface as an open subset of $\mathbb{C}$.

Theorem. Let f be a modular function on $\mathbb{H}$ of weight $2k$. Then

$$f(z)\,\mathrm{d}z^k \in \Omega^k(\mathbb{H})$$

descends to a

$$\omega_f \in \Omega^k(X),$$

meaning

$$(\pi \circ \iota)^* \omega_f = f(z)\,\mathrm{d}z^k,$$

where

$$\mathbb{H} \xhookrightarrow{\iota} \widehat{\mathbb{H}} \xrightarrow{\pi} \widehat{\mathbb{H}}/\mathrm{PSL}_2\,\mathbb{Z}.$$

Moreover, for $p \in X$ with lift $z_p \in \widehat{\mathbb{H}}$, we have

$$\operatorname{ord}_p \omega_f = \begin{cases} \operatorname{ord}_p f & p \neq \pi(i), \pi(\omega_3), \pi(\infty) \\ \frac{1}{2}\operatorname{ord}_i f - \frac{1}{2}k & p = \pi(i) \\ \frac{1}{3}\operatorname{ord}_{\omega_3} f - \frac{2}{3}k & p = \pi(\omega_3) \\ \operatorname{ord}_\infty f - k & p = \pi(\infty) \end{cases},$$

where by $\operatorname{ord}_p f$ I mean $\operatorname{ord}_{z_p} f$ for any lift of p.
□

First note that the expression in the second half of the theorem is independent of choice of lift z_p. This is because $f(gz) = (cz+d)^{2k} f(z)$ with $cz + d \neq 0$, so that

$$\operatorname{ord}_{gz} f = \operatorname{ord}_z(f(g\,\square)) = \operatorname{ord}_z(f(\square)) = \operatorname{ord}_z f.$$

Note also that in the above we make reference to a pullback $(\pi\circ\iota)^*$; this is legit since $\pi\circ\iota\colon \mathbb{H} \longrightarrow \widehat{\mathbb{H}}/G$ is a map of Riemann surfaces. To see why this is true, just recall the diagram from earlier:

$$\begin{tikzcd} U_{z_p} \subset \mathbb{H} \arrow[r, two heads, "\pi"] \arrow[d, "\psi_{z_p}"'] & U_p \subset X \arrow[d, "\phi_p"] \\ \mathbb{C} \arrow[r, "\square^{|\mathrm{Stab}\, z_p|}"] & \mathbb{C} \end{tikzcd}$$

which is so that the map $\pi \circ \iota$ is locally exponentiation, which is holomorphic. Note here we have used the complex structure of $\mathbb{H}$ using the maps ψ_a as charts.

Proof. First we prove the above statements away from ∞. Let $p \neq \infty$, with $m = |\mathrm{Stab}\, z_p|$. Then we have a diagram

$$\begin{tikzcd} U_{z_p} \subset \mathbb{H} \arrow[r, two heads, "\pi"] \arrow[d, "\psi_{z_p}"'] & U_p \subset X \arrow[d, "\phi_p"] \\ \mathbb{C} \arrow[r, "\square^{m}"] & \mathbb{C} \end{tikzcd}$$

where let $z \in U_{z_p}$, $\zeta \in \mathbb{C}$ in the bottom-left, and $w = \zeta^m \in \mathbb{C}$ in the bottom right. Observe

$$\psi_{z_p}^{-1}(\zeta) = \frac{\overline{z}_p \zeta - z_p}{\zeta - 1}.$$

To construct the desired ω_f, we wish to locally describe it in terms of coordinate w.

Compute

$$\begin{aligned} f(z)\,\mathrm{d}z^k &= f(\psi_{z_p}^{-1}\zeta)\,\mathrm{d}(\psi_{z_p}^{-1}\zeta)^k \\ &= f(\psi_{z_p}^{-1}\zeta)(\psi_{z_p}^{-1})'(\zeta)^k\,\mathrm{d}\zeta^k \\ &\coloneqq F(\zeta)\,\mathrm{d}\zeta^k, \end{aligned}$$

where $F(\zeta) \coloneqq f(\psi_{z_p}^{-1}\zeta)(\psi_{z_p}^{-1})'(\zeta)^k$ is a meromorphic function. Since ψ_{z_p} is a biholomorphism, we have

$$\operatorname{ord}_{z_p} f = \operatorname{ord}_0 F.$$

The claim is now that $F(\zeta)\,\mathrm{d}\zeta^k$ is moreover meromorphic in the charts w. To see this, first note for any ω_m a primitive m-th root of unity and g a generator of $\operatorname{Stab} z_p$,

$$\begin{aligned}
F(\omega_m\zeta)\omega_m^k\,\mathrm{d}\zeta^k &= F(\omega_m\zeta)\mathrm{d}(\omega_m\zeta)^k \\
&= f(\psi_{z_p}^{-1}(\omega_m\zeta))\,\mathrm{d}(\psi_{z_p}^{-1}(\omega_m\zeta))^k \\
&= f(g\psi_{z_p}^{-1}\zeta)\,\mathrm{d}(g\psi_{z_p}^{-1}\zeta)^k \\
&= f(gz)\,\mathrm{d}(gz)^k \\
&= f(z)\,\mathrm{d}z^k \\
&= F(\zeta)\,\mathrm{d}\zeta^k,
\end{aligned}$$

so that

$$\omega_m^k F(\omega_m\zeta) = F(\zeta)$$

and therefore

$$(\omega_m\zeta)^k F(\omega_m\zeta) = \zeta^k\omega_m^k F(\omega_m\zeta) = \zeta^k F(\zeta),$$

meaning that

$$\zeta^k F(\zeta)$$

is invariant under coordinate rotation $\zeta \mapsto \omega_m\zeta$. Since ω_m is primitive, it follows that we may write

$$\zeta^k F(\zeta) = G(\zeta^m)$$

for some meromorphic $G(w)$. The one may find

$$\begin{aligned}
F(\zeta)\,\mathrm{d}\zeta^k &= m^{-k}\zeta^{-km+k}F(\zeta)m^k\zeta^{km-k}\,\mathrm{d}\zeta^k \\
&= m^{-k}\zeta^{-km+k}F(\zeta)(m\zeta^{m-1}\,\mathrm{d}\zeta)^k \\
&= m^{-k}\zeta^{-km+k}F(\zeta)\,(\mathrm{d}(z^m))^k \\
&= m^{-k}\zeta^{-km}G(\zeta^m)\,(\mathrm{d}(z^m))^k \\
&= m^{-k}w^{-k}G(w)\,\mathrm{d}w^k,
\end{aligned}$$

so that indeed we obtain a local meromorphic expression

$$m^{-k}w^{-k}G(w)\,\mathrm{d}w^k$$

for a form ω_f in a neighborhood of any $p \neq \infty$.

For the ord part of the theorem, we can check

$$\begin{aligned}
\operatorname{ord}_{z_p} f(z) &= \operatorname{ord}_0 F(\zeta) \\
&= \operatorname{ord}_0 \zeta^{-k}G(\zeta^m) \\
&= -k + \operatorname{ord}_0 G(\zeta^m) \\
&= -k + m\operatorname{ord}_o G(w); \\
\operatorname{ord}_p \omega_f &= \operatorname{ord}_0(m^{-k}w^{-k}G(w)) \\
&= \operatorname{ord}_0(w^{-k}G(w)) \\
&= -k + \operatorname{ord}_0 G(w)
\end{aligned}$$

Solving for $\operatorname{ord}_0 G$ in the first equation fives $\operatorname{ord}_0 G = \frac{\operatorname{ord}_{z_p} f + k}{m}$, and plugging it into the second equation gives

$$\operatorname{ord}_p \omega_f = \frac{1}{m} \operatorname{ord}_{z_p} f - \left(1 - \frac{1}{m}\right) k;$$

in other words,

$$\operatorname{ord}_p \omega_f = \frac{1}{|\operatorname{Stab} z_p|} \operatorname{ord}_{z_p} f - \left(1 - \frac{1}{|\operatorname{Stab} z_p|}\right) k. \tag{$*$}$$

Plugging in $m = 1$ for $p \neq \pi(i), \pi(\omega_3)$ and $m = 2$ for $p = \pi(i)$ and $m = 3$ for $p = \pi(\omega_3)$ (noted in section 1) then gives the claimed.

Secondly we deal with the case $p = \infty$. Recall that our complex structure for X sent

$$\begin{aligned} \phi_\infty \colon U_\infty = U_{z_\infty}/\langle T\rangle &\longrightarrow \mathbb{C} \\ \pi(z) &\longmapsto e^{2\pi i z}. \end{aligned}$$

Letting $q = e^{2\pi i z}$ be this local coordinate with $f(z) = \widehat{f}(q)$, we have $\mathrm{d}q = 2\pi i e^{2\pi i z}\,\mathrm{d}z$, or

$$\mathrm{d}z = \frac{1}{2\pi i q}\,\mathrm{d}q;$$

this means

$$f(z)\,\mathrm{d}z^k = \widehat{f}(q) \frac{1}{(2\pi i q)^k}\,\mathrm{d}q^k.$$

By assumption (this finally comes in) we have f is a modular function, meaning $\widehat{f}(q)$ meromorphically extends to $q = 0$. Hence we obtain a local meromorphic expression

$$\widehat{f}(q) \frac{1}{(2\pi i q)^k}$$

for a form ω_f in a neighborhood of $p = \infty$.

For the ord part of the theorem, simply observe

$$\operatorname{ord}_\infty \omega_f = \operatorname{ord}_0 \widehat{f}(q) \frac{1}{(2\pi i q)^k} = -k + \operatorname{ord}_\infty f,$$

as claimed. ■

5. The Valence Formula

The valence formula states

Theorem (Valence)**.** For f not identically zero a modular function of weight $2k$,

$$\operatorname{ord}_\infty f + \frac{1}{2} \operatorname{ord}_i f + \frac{1}{3} \operatorname{ord}_{\omega_3} f + \sum_{\substack{p \in \widehat{\mathbb{H}}/\mathrm{PSL}_2 \mathbb{Z} \\ \pi^{-1} p \not\ni \infty, i, \omega_3}} \operatorname{ord}_p f = \frac{k}{6},$$

where the last sum is over all the representatives of $X = \widehat{\mathbb{H}}/\mathrm{PSL}_2 \mathbb{Z}$ except those descending from ∞, i, ω_3.
□

Just as in the last section, the modularity of f means that the order is independent of choice of lift.

To prove this theorem, we will appeal to a version of Riemann-Roch, as stated in Silverman:

Theorem (Riemann-Roch)**.** If X is a compact Riemann surface of genus g, with $\omega \in \Omega^k(X)$, then
$$\deg(\operatorname{div}\omega) = k(2g-2).$$
□

The proof of this theorem is very difficult, and we assume it for this note.

Armed with this powerful fact, we then prove valence:

Proof of Valence. Let ω_f be the corresponding symmetric meromorphic k-form on $X = \widehat{\mathbb{H}}/G$ to f, given by our work in section 4. Recall that X has genus $g = 0$; hence, by Riemann-Roch,
$$\deg(\operatorname{div}\omega_f) = -2k.$$
On the other hand, explicitly computing the sum defining this degree gives
$$\begin{aligned}\deg(\operatorname{div}\omega_f) &= \sum_{p\in X} \operatorname{ord}_p \omega_f \\ &= (\operatorname{ord}_\infty f - k) + \left(\frac{1}{2}\operatorname{ord}_i f - \frac{1}{2}k\right) + \left(\frac{1}{3}\operatorname{ord}_{\omega_3} f - \frac{2}{3}k\right) + \sum_{\substack{p\in\widehat{\mathbb{H}}/\mathrm{PSL}_2\mathbb{Z} \\ \pi^{-1}p\not\ni\infty,i,\omega_3}} \operatorname{ord}_p f \\ &= -\frac{13}{6}k + \operatorname{ord}_\infty f + \frac{1}{2}\operatorname{ord}_i f + \frac{1}{3}\operatorname{ord}_{\omega_3} f + \sum_{\substack{p\in\widehat{\mathbb{H}}/\mathrm{PSL}_2\mathbb{Z} \\ \pi^{-1}p\not\ni\infty,i,\omega_3}} \operatorname{ord}_p f;\end{aligned}$$
putting the two together gives
$$\frac{k}{6} = \operatorname{ord}_\infty f + \frac{1}{2}\operatorname{ord}_i f + \frac{1}{3}\operatorname{ord}_{\omega_3} f + \sum_{\substack{p\in\widehat{\mathbb{H}}/\mathrm{PSL}_2\mathbb{Z} \\ \pi^{-1}p\not\ni\infty,i,\omega_3}} \operatorname{ord}_p f,$$
thanks to the powerful fact that $\frac{13}{6} - 2 = \frac{1}{6}$. This concludes. ■

Alternatively, one might give a more basic proof (without appealing to Riemann-Roch) of this fact as follows:

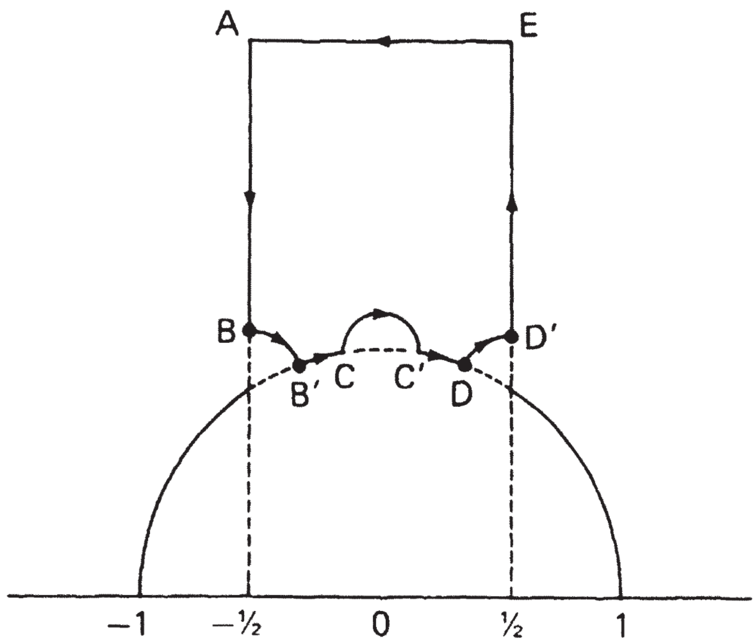
Basic Proof of Valence. Firstly we claim the sum is well-defined since it is finite. In fact, we claim f has only a finite number of zeros and poles in $\mathbb{H}/G$. We check this below:

* Since $\widehat{f}$ is meromorphic on $|q| < 1$, there is $r > 0$ such that $\widehat{f}$ has no zero or pole on the annulus[9] Anus_0^r. Translating this from q to z, $0 < |q| = e^{-2\pi \operatorname{im} z} < r = e^{\log r} \iff$
$$\operatorname{im} z > \frac{1}{2\pi}\log\frac{1}{r}$$
is the region on which f has no zero or pole (note this doesn't include infinity since we require $0 < |q|$). But the rest of $\mathbb{H}/G$, $\mathbb{H}/G \cap \left\{\operatorname{im} z \le \frac{1}{2\pi}\log\frac{1}{r}\right\}$, is compact, so we can apply the identity theorem to conclude f only has finitely many zeros and poles on $\mathbb{H}/G \cap \left\{\operatorname{im} z \le \frac{1}{2\pi}\log\frac{1}{r}\right\}$ and therefore on all of $\mathbb{H}/G$ (there can be at most one zero/pole at infinity, but in any case infinity is not taken into consideration in $\sum^\times$).

To show the formula, we will consider
$$\frac{1}{2\pi i}\oint_{\partial\Omega} \frac{\mathrm{d}f}{f}$$

[9] Recall from 55b that Anus stands for ANnulUS.

where $\partial\Omega$ (depending on ε) is the below curve where each arc BB', CC', DD' is of a circle of radius ε



where the top horizontal edge is high enough to contain all finite zeros/poles of f in $\mathbb{H}/G$ (see argument above). For now suppose there are no zeros or poles of f on the boundary of $\mathbb{H}/G$ except possibly $i, \omega_3, -\overline{\omega}_3$.

Then the argument principle states

$$\frac{1}{2\pi i}\oint_{\partial\Omega}\frac{\mathrm{d}f}{f} = \sum_{p\in\mathbb{H}/G}^{\times} \operatorname{ord}_p f.$$

We can however also calculate this explicitly.

* Over EA, note $q = e^{2\pi i z}$ transforms EA to a small circle $-B_r(0)$ centered at $q = 0$ with negative orientation (in fact as $\operatorname{re} z$ goes from $1/2$ to $-1/2$ the angle goes from π to $-\pi$). As per our argument above, this circle encloses no zero or pole of $\widehat{f}$ except possible at $q = 0$. So

$$\frac{1}{2\pi i}\int_E^A \frac{\mathrm{d}f}{f} = \frac{1}{2\pi i}\oint_{-B_r(0)}\frac{\mathrm{d}f}{f} = -\operatorname{ord}_\infty f.$$

* Along AB and $D'E$, recall $f(z+1) = f(z)$; hence the integrals perfectly cancel out:

$$\frac{1}{2\pi i}\int_A^B \frac{\mathrm{d}f}{f} + \frac{1}{2\pi i}\int_{D'}^E \frac{\mathrm{d}f}{f} = 0.$$

* Consider the arc BB' as a part of the circle centered at ω_3 of radius ε. If we were to integrate over this whole circle, we would pick up $-\operatorname{ord}_{\omega_3} f$ as $\varepsilon \to 0$ (negative sign due to orientation). As $\varepsilon \to 0$, the arc BB' cuts an angle of $\pi/6$. Recall from complex analysis that in general for a function meromorphic in a neighborhood of p, the integral over a circular arc γ centered at p of radius ε and angle α has

$$\lim_{\varepsilon\to 0}\frac{1}{2\pi i}\int_{\gamma_\varepsilon} f\,\mathrm{d}\zeta = \frac{\alpha}{2\pi}\operatorname{res}_p(f\,\mathrm{d}\zeta).$$

We conclude

$$\lim_{\varepsilon\to 0}\frac{1}{2\pi i}\int_B^{B'}\frac{\mathrm{d}f}{f} = -\frac{1}{6}\operatorname{ord}_{\omega_3} f.$$

* Similarly, along CC' we have

$$\lim_{\varepsilon\to 0}\frac{1}{2\pi i}\int_C^{C'}\frac{\mathrm{d}f}{f} = -\frac{1}{2}\operatorname{ord}_i f$$

and

* along DD' we have

$$\lim_{\varepsilon\to 0}\frac{1}{2\pi i}\int_D^{D'}\frac{\mathrm{d}f}{f} = -\frac{1}{6}\operatorname{ord}_{1+\omega_3} f = -\frac{1}{6}\operatorname{ord}_{\omega_3} f.$$

* Lastly, the transformation $Sz = -\frac{1}{z}$ maps $B'C$ to DC' in that order (orientation-reversing). Recalling $f(Sz) = z^{2k}f(z)$, we can see

$$\begin{aligned} f(Sz) &= z^{2k}f(z) \\ \mathrm{d}f(Sz) &= 2kz^{2k-1}f(z)\,\mathrm{d}z + z^{2k}\,\mathrm{d}f(z) \\ \frac{\mathrm{d}f(Sz)}{z^{2k}f(z)} &= \frac{2k}{z}\,\mathrm{d}z + \frac{\mathrm{d}f(z)}{f(z)}, \end{aligned}$$

i.e.

$$\frac{\mathrm{d}f(Sz)}{f(Sz)} = 2k\frac{\mathrm{d}z}{z} + \frac{\mathrm{d}f(z)}{f(z)}.$$

This gives

$$\begin{aligned} \frac{1}{2\pi i}\int_{B'}^{C}\frac{\mathrm{d}f}{f} + \frac{1}{2\pi i}\int_{C'}^{D}\frac{\mathrm{d}f}{f} &= \frac{1}{2\pi i}\int_{B'}^{C}\frac{\mathrm{d}f(z)}{f(z)} + \frac{1}{2\pi i}\int_{S^{-1}(C'D)}\frac{\mathrm{d}f(Sz)}{f(Sz)} \\ &= \frac{1}{2\pi i}\int_{B'}^{C}\frac{\mathrm{d}f(z)}{f(z)} + \frac{1}{2\pi i}\int_{-B'C}\frac{\mathrm{d}f(Sz)}{f(Sz)} \\ &= \frac{1}{2\pi i}\int_{B'}^{C}\left(\frac{\mathrm{d}f(z)}{f(z)} - \frac{\mathrm{d}f(Sz)}{f(Sz)}\right) \\ &= \frac{1}{2\pi i}\int_{B'}^{C} -2k\frac{\mathrm{d}z}{z} \\ &\xrightarrow[\varepsilon\to 0]{} -2k\cdot -\frac{1}{12} \\ &= \frac{k}{6}, \end{aligned}$$

where we note as $\varepsilon \to 0$ the arc $B'C$ cuts out an angle of $\pi/12$ along the circle $\partial B_1(0)$. We conclude

$$\lim_{\varepsilon\to 0}\frac{1}{2\pi i}\int_{B'}^{C}\frac{\mathrm{d}f}{f} + \frac{1}{2\pi i}\int_{C'}^{D}\frac{\mathrm{d}f}{f} = \frac{k}{6}.$$

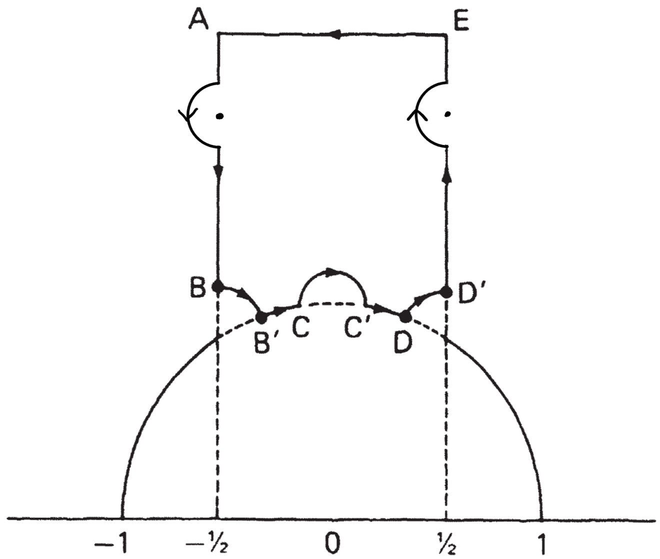
Putting these things together and taking $\varepsilon \to 0$, we find

$$-\operatorname{ord}_\infty f - \frac{1}{6}\operatorname{ord}_{\omega_3} f - \frac{1}{2}\operatorname{ord}_i f - \frac{1}{6}\operatorname{ord}_{\omega_3} f + \frac{k}{6} = \sum_{p\in\mathbb{H}/G}^{\times}\operatorname{ord}_p f,$$

as desired.

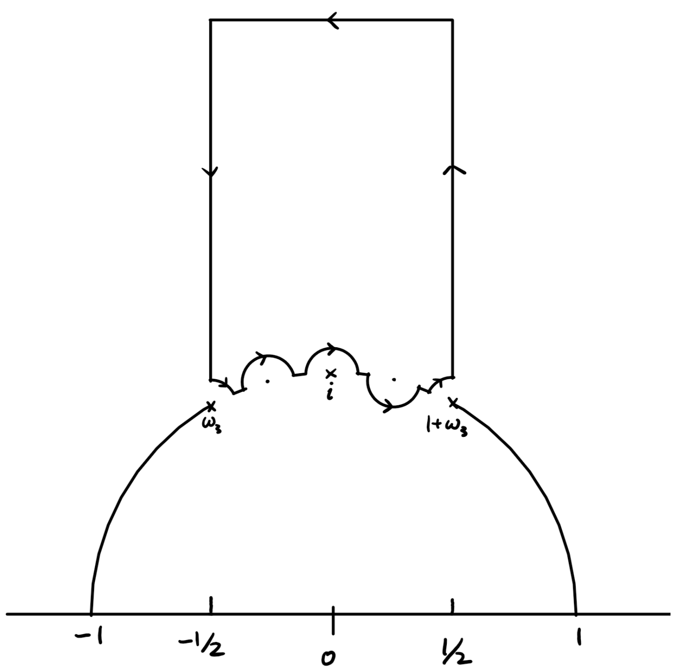
In the case that there are other zeros/poles on the boundary of $\mathbb{H}/G$, we will modify our curve in the following ways:

* If there is a zero/pole on the vertical boundaries $\operatorname{re} z = \pm\frac{1}{2}$, modify our curve to be as follows:



This does not change our computations above, as the translation $Tz = z + 1$ still maps AB to ED', and again the opposite orientations cause the integrals to cancel.

* If there is a zero/pole on the lower circular boundary $|z| = 1$, modify our curve to be as follows:



again so that the new arc on the left is mapped to the new arc on the right by $Sz = -\frac{1}{z}$ with negative orientation, so that the computation above acquiring $\frac{k}{6}$ still carries through.

This concludes. ∎

6. The Dimension Formula

We will now apply the valence formula to deduce some facts about dimension.

6.1. **Examples of Modular Forms.** An important example of modular forms come from Eisenstein series, which are defined by

Definition. The k-th Eisenstein series is

$$G_k(z) = \sum_{m,n}^{\times} \frac{1}{(mz+n)^{2k}}$$

for $z \in \mathbb{H}$, where $\sum^{\times}$ indicates that the sum runs over all pairs of integers except $(0,0)$.

It is a fact that

Fact. For $k \geq 2$, G_k is a modular form of weight $2k$ and $G_k(\infty) = 2\zeta(2k)$.
□

Let us convince ourselves of this fact.

Convince. Firstly, for $k \geq 2$, the series defining G_k converges absolutely since the number of pairs (m,n) for which

$$N \leq |mz+n| \leq N+1$$

is $\mathscr{O}(N)$, so the sum $\sum_{m,n}^{\times} \frac{1}{(mz+n)^{2k}}$ looks like $\sum_N \frac{1}{N^{2k-1}}$, which converges for $k \geq 2$.

Secondly, G_k satisfies

$$G_k(z+1) = \sum_{m,n}^{\times} \frac{1}{(mz+m+n)^{2k}} = \sum_{m,(m+n)}^{\times} \frac{1}{(mz+(m+n))^{2k}} = G_k(z),$$

where as (m,n) avoids $(0,0)$ so does $(m, m+n)$. G_k also satisfies

$$G_k\left(-\frac{1}{z}\right) = \sum_{m,n}^{\times} \frac{1}{(-m/z+n)^{2k}} = \sum_{m,n}^{\times} z^{2k} \frac{1}{(-m+nz)^{2k}} = z^{2k} \sum_{n,-m}^{\times} \frac{1}{(nz-m)^{2k}} = z^{2k} G_k(z),$$

where as (m,n) avoids $(0,0)$ so does $(n,-m)$.

Thirdly, $G_k \in \mathscr{C}^+(\mathbb{H})$ for the following reason: by the "secondly" we have checked $G_k(gz) = (cz+d)^{2k} G_k(z)$ for all $g \in \mathrm{PSL}_2\,\mathbb{Z}$, and since the fundamental domains partition $\mathbb{H}$, we can use $G_k(gz) = (cz+d)^{2k} G_k(z)$ to go from any fundamental domain to any other, where $(cz+d)^{\pm 2k}$ is holomorphic on $\mathbb{H}$; hence it suffices to see that G_k is holomorphic on a single $\mathbb{H}/G$, say the D from earlier. But here note

$$|mz+n|^2 = m^2 z\bar{z} + 2mn \operatorname{re} z + n^2 \geq m^2 - mn + n^2 = |m\omega_3 - n|^2,$$

where equality is achieved at the left-bottom-most corner of D, ω_3. Then each term has

$$\left| \frac{1}{(mz+n)^{2k}} \right| \leq \frac{1}{|m\omega_3 - n|^2}$$

on all of D; but by "firstly" the sum

$$\sum_{m,n}^{\times} \frac{1}{|m\omega_3 - n|^2}$$

is convergent, so by the Weierstrass M-test we conclude $G_k(z)$ converges uniformly and absolutely on D, so in particular uniformly on any compact subset of D; and since each term $\frac{1}{(mz+n)^{2k}}$ is holomorphic on $\mathbb{H}$, we conclude the sum $G_k(z)$ is also holomorphic on D, and therefore on all of $\mathbb{H}$ by using G-invariance. ■

In seeking the value $G_k(\infty)$, note that as $q \to 0$, $\operatorname{im} z \to \infty$. Since the sum for G_k converges uniformly, we may take this limit termwise to find

$$\lim_{z \to i\infty} G_k(z) = \sum_{m,n}^{\times} \lim_{z \to i\infty} \frac{1}{(mz+n)^{2k}} = \sum_{n \neq 0} \frac{1}{n^{2k}} = 2\zeta(2k),$$

as advertised. Recall that ζ has no zeros beyond, say, $s \geq 4$. Hence the $G_k(z)$ do not vanish at infinity and are not cusp forms.

A construct from the Eisenstein series which will be useful later is as follows. Let

$$g_2 = 60G_2, \quad g_3 = 140G_3,$$

so that

$$g_2(\infty) = 120\zeta(4) = \frac{4\pi^4}{3}, \quad g_3(\infty) = 280\zeta(6) = \frac{8\pi^6}{27}.$$

The construct

Definition.

$$\Delta \coloneqq g_2^3 - 27g_3^2$$

then has the convenient property of being a modular form of weight 12 such that

$$\Delta(\infty) = 0,$$

i.e. a cusp form of weight 12.

6.2. **Space of Modular Forms.** We now proceed to discussing dimension formulae. But first let's describe what we're taking the dimension of.

Definition. Let $M_k = M_k(\mathrm{PSL}_2\,\mathbb{Z})$ denote the $\mathbb{C}$-vector space of modular forms of weight $2k$, and let $M_k^0 = M_k^0(\mathrm{PSL}_2\,\mathbb{Z})$ denote that of cusp forms of weight $2k$.

From our comments at the end of section 3 it is clear that M_k does in fact admit a vector space structure (scalar multiplication obviously preserves modular-ness). We also saw earlier that, for $k \geq 2$, $G_k(\infty) = 2\zeta(2k) \neq 0$, so given any modular form f we can write it as

$$f = \left(f - \frac{f(\infty)}{2\zeta(2k)}G_k\right) + \frac{f(\infty)}{2\zeta(2k)}G_k,$$

where the former vanishes at infinity. Hence

Fact.

$$M_k = M_k^0 \oplus \mathbb{C}G_k$$

for $k \geq 2$.

□

In particular this means

$$\dim M_k = \dim M_k^0 + 1$$

for $k \geq 2$.

In general, noting that M_k^0 is the kernel of the map $M_k \to \mathbb{C}$ via $f \mapsto f(\infty)$, we get the short exact sequence

$$0 \longrightarrow M_k^0 \longrightarrow M_k \longrightarrow \mathbb{C} \longrightarrow 0,$$

where the last map is surjective since we can scale f provided $M_k \neq 0$, so that there is something to scale at all. This immediately gives

$$\dim M_k / M_k^0 \leq 1$$

(strict equality if $M_k = 0$) and in fact

$$\dim M_k = \dim M_k^0 + 1$$

provided $M_k \neq 0$.

First we do the small cases:

Theorem.

- $M_k = 0$ for $k < 0$ and $k = 1$.
- For $k = 0, 2, 3, 4, 5$, M_k has dimension 1 with basis G_k, i.e.

$$\begin{aligned} M_0 &= \mathbb{C}, \\ M_2 &= \mathbb{C}G_2, \\ M_3 &= \mathbb{C}G_3, \\ M_4 &= \mathbb{C}G_4, \\ M_5 &= \mathbb{C}G_5. \end{aligned}$$

- Multiplication by Δ gives an isomorphism

$$\times\Delta \colon M_{k-6} \xrightarrow{\sim} M_k^0.$$

□

Here recall

$$\Delta = g_2^3 - 27 g_3^2$$

satisfies

$$\Delta(\infty) = 0.$$

Proof. Let $f \in M_k$ be a nonzero element. The valence formula

$$\operatorname{ord}_\infty f + \frac{1}{2} \operatorname{ord}_i f + \frac{1}{3} \operatorname{ord}_{\omega_3} f + \sum_{\substack{p \in \widehat{\mathbb{H}}/\mathrm{PSL}_2 \mathbb{Z} \\ \pi^{-1} p \not\ni \infty, i, \omega_3}} \operatorname{ord}_p f = \frac{k}{6}$$

has everything on the left hand side nonnegative (since f is holomorphic everywhere by definition). Hence we must have $k \geq 0$ in order for f to exist, so $M_{<0} = 0$. Also since $k = 1$ has $\frac{1}{6}$ on the right which cannot be written as $a + b/2 + c/3$ for integers a, b, c, we also conclude $M_1 = 0$. This gives the first bullet.

For $k = 2$, the only possibility is if $\operatorname{ord}_{\omega_3} f = 1$ and f vanishes nowhere else on $\mathbb{H}/G$. Similarly, for $k = 3$, the only possibility is if $\operatorname{ord}_i f = 1$ and f vanishes nowhere else on $\mathbb{H}/G$. Then, plugging in for example $\Delta(i) = g_2(i)^3 - 27\cancel{g_3(i)^2} = g_2(i)^3$, since $g_2 \in M_2$ cannot vanish anywhere in $\mathbb{H}/G$ other than at ω_3, we conclude $\Delta(i) \neq 0$ and so $\Delta \neq 0$ is a nonzero function. Hence we can apply the valence formula to $\Delta \in M_6$: we already saw $\Delta(\infty) = 0$, so $\operatorname{ord}_\infty \Delta \geq 1$; but the right hand side is already $\frac{k=6}{6} = 1$, so we conclude

$$\operatorname{ord}_\infty \Delta = 1$$

and Δ vanishes nowhere else on $\mathbb{H}/G$. Since this means Δ has a simple zero at infinity, for any element $f \in M_k^0$ we have $\frac{f}{\Delta}$ is of weight $2(k-6)$ and is holomorphic everywhere (f vanishes at infinity, and Δ has only a simple zero, so there is no pole at infinity), as can be seen by

$$\operatorname{ord}_p \frac{f}{\Delta} = \operatorname{ord}_p f - \operatorname{ord}_p \Delta = \begin{cases} \operatorname{ord}_p f & p \neq \infty \\ \operatorname{ord}_p f - 1 & p = \infty \end{cases},$$

where $\operatorname{ord}_\infty f \geq 1$. Hence $\frac{f}{\Delta} \in M_{k-6}$ is a preimage of any $f \in M_k^0$, so $\times\Delta$ is a surjective map. Since it is already clearly linear and injective (no product of two holomorphic nonzero functions will be zero), we conclude

$$\times\Delta\colon M_{k-6} \xrightarrow{\sim} M_k^0$$

is an isomorphism as desired. This is the third bullet.

Lastly, for $k = 0, 2, 3, 4, 5$, we have $M_{k-6} = 0$, so by the above isomorphism we conclude $M_k^0 = 0$ also, which by our earlier fact relating M_k and M_k^0 means, for $k = 2, 3, 4, 5$, $\dim M_k = 1$; since we saw earlier that $G_k \in M_k$, we conclude G_k are bases for M_k for $k = 2, 3, 4, 5$. As for $k = 0$, we get $\dim M_0/M_0^0 = \dim M_0 \leq 1$; but the constant function $1 \in M_0$ clearly, so we conclude $M_0 = \mathbb{C}$. This concludes the second bullet and the theorem. ■

It is now a direct corollary that

Theorem. The dimensions of M_k and M_k^0 for $k \geq 0$ are

$$\dim M_k = \begin{cases} \left\lfloor \frac{k}{6} \right\rfloor + 1 & k \not\equiv 1 \ (6) \\ \left\lfloor \frac{k}{6} \right\rfloor & k \equiv 1 \ (6) \end{cases},$$

$$\dim M_k^0 = \begin{cases} \left\lfloor \frac{k}{6} \right\rfloor & k \not\equiv 1 \ (6) \\ \left\lfloor \frac{k}{6} \right\rfloor - 1 & k \equiv 1 \ (6) \end{cases}.$$

□

This follows directly by induction. By the isomorphism $\times\Delta$, we have

$$\dim M_k = \dim M_{k+6}^0 = \dim M_{k+6} - 1,$$

and the conclusion is clear.

7. References

Serre, Jean-Pierre. "A Course in Arithmetic". 1973.

Silverman, Joseph. "Advanced Topics in the Arithmetic of Elliptic Curves". 1994.

Murty, Maruti. "Problems in the Theory of Modular Forms". 2016.