Modular Forms

Sarah Zerbes

Spring semester 2021/22

Contents

0 Prologue

Example 0.0.1. Let $z \in \mathbb{C}$, $\Im(z) > 0$. Let $q = e^{2\pi i z}$ and define **Ramanujan's tau function**

$$\Delta(z) = q \cdot \prod_{n \in \mathbb{N}} (1 - q^n)^{24}.$$

This is one of the simplest examples of a modular form. Note that we can "multiply out" the product above which leads us to

$$\Delta(z) = \sum_{n \in \mathbb{N}} \tau(n) q^n$$

for some integers $\tau(n)$.

Facts 0.0.2.

(1) Known to Weierstrass, 1850:

$$\Delta(z) = z^{-12} \cdot \Delta\left(-\frac{1}{z}\right)$$

(2) Ramanujan proved in 1916 that the integers $\tau(n)$ satisfy the equation

$$\tau(n) = \sum_{d \mid n} d^{11} \mod 691.$$

(3) Ramanujan also conjectured $\tau(nm) = \tau(n)\tau(m)$ for n, m coprime. This was proved by Mordell in 1917.

(4) In 1972 Swinnerton-Dyer proved $\tau(n)$ satisfies congruences like (2) modulo 2, 3, 5, 7, 23 and 691 but no other primes.

(5) Ramanujan conjectured in 1916 for p prime holds $|\tau(p)| < 2\, p^{11/2}$. This was proved in 1974 by Deligne.

(6) The quantity

$$\frac{\tau(p)}{2p^{11/2}} \in [-1, 1]$$

is distributed in the interval $[-1, 1]$ with density function proportional to $\sqrt{1 - x^2}$. This was conjectured by Sato and Tate (1960s) and proved by Barnet-Lamb, Geraghty, Harris and Taylor in 2009 using Bau Chau Ngo's *Fundamental Lemma* which got Ngo the 2010 Fields Medal.

Example 0.0.3. We now consider another modular form

$$
\begin{aligned}
f(z) &= q\prod_{n=1}^{\infty}(1-q^n)^2(1-q^{11n})^2 \\
&= q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 + \dots \\
&= \sum_{n=1}^{\infty} a(n)q^n \qquad ' \qquad \text{with } a(n) \in \mathbb{N}
\end{aligned}
$$

We will later prove the following results:

Theorem.

1. *We have $a(mn) = a(m)(n)$ for all $m, n \geq 1$ with $(m, n) = 1$.*
2. *We have $|a(p)| \leq 2\sqrt{p}$ for all primes p.*

It turns out that this modular form is closely related to the elliptic curve

$$E : Y^2 + Y = X^3 - X^2 - 10X - 20.$$

For p prime, denote by $N(p)$ the number of points on the elliptic curve in $\mathbb{F}_p$. It is easy to see heuristically tat $N(p) \simeq p$.

Theorem. *(Hasse) We have*

$$|p - N(p)| \leq 2\sqrt{p}.$$

The theory of modular forms allows one to prove that the elliptic curve E and the modular form f 'correspond' to each other in the following sense:

Theorem. *For all primes p, we have*

$$a(p) = p - N(p).$$

In particular, using the properties of the modular form f, we can easily calculate the quantity $N(p)$ for all p, so f 'knows' about the behaviour of the elliptic curve over $\mathbb{F}_p$. We say that the elliptic curve E is **modular**. It is generally not too difficult to attach an elliptic curve to a modular form (this is called "Eichler–Shimura"); however, it is very difficult indeed to reverse this process, and this is the basis of Andrew Wiles' work on Fermat's Last Theorem. The proof of this result was later completed by Breuil–Conrad–Diamond–Taylor. I will talk a bit more about this when we discuss L-functions of modular forms.

1 The modular group

1.1 The upper half-plane

Definition 1.1.1. Let $\mathcal{H} = \{z \in \mathbb{C} \colon \Im(z) > 0\}$ the **upper half-plane**.

Proposition 1.1.2. *The* ***special linear group*** $\mathrm{SL}_2(\mathbb{R}) = \{A \in \mathrm{GL}_2(\mathbb{R}) \colon \det(A) = 1\}$ *acts on* $\mathcal{H}$ *via*

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

Proof. For $z \in \mathcal{H}$ is $\Im(z) > 0$ and either c or d is nonzero, so $cz + d \neq 0$. Moreover

$$\Im\left(\frac{az+b}{cz+d}\right) = \frac{1}{|cz+d|^2}\,\Im\left((az+b)(c\overline{z}+d)\right).$$

Say $z = x + iy$ for $x, y \in \mathbb{R}$.

$$\begin{aligned}\Im\left(\frac{az+b}{cz+d}\right) &= \frac{1}{|cz+d|^2}\,\Im\Big(\underbrace{(ax+b)(cx+d)+acy^2}_{\in\mathbb{R}} + i\underbrace{(ad-bc)}_{=1}y\Big) \\ &= \frac{1}{|cz+d|^2}\,\Im(z) \;>\; 0\end{aligned}$$

Therefore $\frac{az+b}{cz+d} \in \mathcal{H}$ for any $z \in \mathcal{H}$, $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$.

Also it is easy to check that $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)z = z$ and $A(Bz) = (AB)z$ for any $z \in \mathcal{H}$ and for any $A, B \in \mathrm{SL}_2(\mathbb{R})$. Thus $\mathrm{SL}_2(\mathbb{R})$ acts on $\mathcal{H}$. □

Note 1.1.3. The matrix $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ acts trivially on $\mathcal{H}$, so the action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathcal{H}$ factors through the quotient $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/(\pm 1)$, the **projective special linear group**.

Definition 1.1.4. The *automorphy factor* is the function

$$\begin{aligned} j \,:\, \mathrm{SL}_2(\mathbb{R}) \times \mathcal{H} &\to \mathbb{C}, \\ (g, z) &\mapsto cz + d \qquad\qquad \text{for } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\end{aligned}$$

Proposition 1.1.5. *For any* $k \in \mathbb{Z}$*, we can define a right action of* $\mathrm{SL}_2(\mathbb{R})$ *on the set of holomorphic functions* $\mathcal{H} \to \mathbb{C}$ *given by*

$$(f|_k g)(z) := j(g,z)^{-k}\; f(gz)$$

where $f \colon \mathcal{H} \to \mathbb{C}$ *holomorphic,* $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{R})$*. We will call this the* ***weight k action****.*

Proof. Firstly we need to show that $f|_k g$ is a well-defined holomorphic function $\mathcal{H} \to \mathbb{C}$. But this is obvious since $cz + d \neq 0$ and $gz \in \mathcal{H}$ for all $z \in \mathcal{H}$. Clearly also the equation $f|_k 1 = f$ holds. Therefore it remains to show $(f|_k g)|_k h = f|_k(gh)$ for arbitrary $g, h \in \mathrm{SL}_2(\mathbb{R})$. The left hand side of the equation can be rewritten as

$$\begin{aligned}(f|_k g)|_k h &= j(h,z)^{-k}\left((f|_k g)(hz)\right)\\ &= j(h,z)^{-k} j(g,hz)^{-k} f(g(hz))\end{aligned}$$

and the right hand side results in

$$f|_k(gh) = j(gh,z)^{-k} f((gh)z).$$

We already know $(gh)z = g(hz)$. So it remains to show $j(gh,z) = j(h,z)j(g,hz)$. This is the so called **cocycle relation** and can be checked easily. □

1.2 The modular group

Definition 1.2.1. The **modular group** is the group

$$\mathrm{SL}_2(\mathbb{Z}) = \left\{ A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} ;\, a,b,c,d \in \mathbb{Z},\ \det(A) = 1 \right\}.$$

The **projective modular group** is $\mathrm{PSL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z})/(\pm 1)$.

Theorem 1.2.2. *(a) The group $\mathrm{SL}_2(\mathbb{Z})$ is generated by $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$.*

(b) Every orbit of $\mathrm{SL}_2(\mathbb{Z})$ acting on $\mathcal{H}$ contains a point of the set D defined by

$$D = \left\{ z \in \mathcal{H}\colon\ -\frac{1}{2} \leq \Re(z) \leq \frac{1}{2} \text{ and } |z| \geq 1 \right\}.$$

(c) If $z \in D$ and $gz \in D$ for some $g \in \mathrm{SL}_2(\mathbb{Z})$, then either $g = \pm 1$ and $gz = z$ or z lies on the boundary of D.

(d) The stabilizer of $z \in \mathcal{H}$ in $\mathrm{PSL}_2(\mathbb{Z})$ is trivial unless z is in the orbit of i or in the orbit of $\rho = e^{2\pi i/3}$.

Proof. We will prove all of these statements in 4 steps using a very elegant argument of Serre. Let $G = \mathrm{SL}_2(\mathbb{Z})$ and $G' = \langle S, T\rangle \leq G$.

Step 1. Every G' orbit in $\mathcal{H}$ contains a point of D.

Proof of Step 1. Let $z \in \mathcal{H}$. Since $|cz+d| \geq |c\,\Im(z)|$ and $|cz+d| \geq |c\,\Re(z)+d|$ there exist only finitely many $(c,d) \in \mathbb{Z}^2$ such that $|cz+d| < 1$. Recall $\Im(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)z) = |cz+d|^{-2}\,\Im(z)$. This implies there are only finitely many $g \in G'$ such that $\Im(gz) > \Im(z)$. So the G' orbit of z contains a point of maximal imaginary part. Let this point be z.

We can assume $\Re(z) \in [-\frac{1}{2}, \frac{1}{2}]$ since $Tz = z+1$. Moreover $\Im(Sz) = |z|^{-2}\,\Im(z)$. But z is a point of maximal imaginary part in the orbit of G', so we get $|z|^{-2}\,\Im(z) \leq \Im(z)$ implying $|z| \geq 1$. Thus $z \in D$. Clearly this proves part (b) of the theorem. □

Step 2. If $z \in D$ and $gz \in D$, where $g \in G$, then one of the following holds:

1. $g = \pm \mathrm{Id}$
2. $g = \pm S$ and $|z| = 1$
3. $g = \pm T$ and $\Re(z) = -\frac{1}{2}$, or $g = \pm T^{-1}$ and $\Re(z) = \frac{1}{2}$
4. $g = \pm ST = \pm\left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right)$ or $g = \pm T^{-1}S = \pm\left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ or $g = \pm ST^{-1}S = \pm\left(\begin{smallmatrix} -1 & 0 \\ -1 & -1 \end{smallmatrix}\right)$ and $z = \rho$
5. $g = \pm TS = \pm\left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ or $g = \pm ST^{-1} = \pm\left(\begin{smallmatrix} 0 & -1 \\ 1 & -1 \end{smallmatrix}\right)$ or $g = \pm STS = \pm\left(\begin{smallmatrix} -1 & 0 \\ 1 & -1 \end{smallmatrix}\right)$ and $z = \rho + 1$

Proof of Step 2. Let $z \in D$ and $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G$ such that $z' = gz \in D$. Being free to replace g by g^{-1} and z by z' we can assume that $\Im(z') \geq \Im(z)$. Again recalling $\Im(gz) = |cz + d|^{-2}\, \Im(z)$ we gain $|cz + d| \leq 1$. Furthermore we have

$$|cz + d| \geq |c|\, \Im(z) \geq |c|\, \Im(\rho) = \frac{\sqrt{3}}{2}|c|.$$

Thus $|c| \leq 2/\sqrt{3} < 2$. As $c \in \mathbb{Z}$ we get $c = 0$ or $c = \pm 1$.

- Let $c = 0$. Since $1 \geq |cz + d| = |d|$ we have $d = 0$ or $d = \pm 1$. But $c = d = 0$ is impossible. So $d = \pm 1$ and hence $a = \pm 1$. Therefore $g = \left(\begin{smallmatrix} \pm 1 & b \\ 0 & \pm 1 \end{smallmatrix}\right)$ is the translation by b. But since
$$\Re(z),\, \Re(gz) \in \left[-\frac{1}{2}, \frac{1}{2}\right],$$
this implies that $b = 0$ or $b = \pm 1$. So either $g = \pm \mathrm{Id}$ (case 1) or $g = \pm T$ and $\Re(z) = -\frac{1}{2}$ or $g = \pm T^{-1}$ and $\Re(z) = \frac{1}{2}$.

- Let $c = 1$. Assuming $|d| \geq 2$ leads to the following contradiction:
$$1 \geq |cz + d| = |z + d| \geq |d| - \Re(z) \geq |d| - \frac{1}{2} \geq \frac{3}{2}$$
Thus we have $d = 0$ or $d = \pm 1$.

  Let $d = 0$. Then $1 \geq |cz + d| = |z|$. On the other hand $|z| \geq 1$ as $z \in D$ and therefore $|z| = 1$ (cases 2, 4 or 5 – exercise sheet 1).

  Let $d = 1$. Then $1 \geq |z + 1|$. This is only possible for $z \in D$ if $z = \rho$ (exercise). Since $a - b = 1$, we deduce that wither $(a, b) = (1, 0)$ or $(a, b) = (0, -1)$ (case 4).

  Analogue $d = -1$ implies $z = \rho + 1$ (case 5).

- The case $c = -1$ is analogous to the case $c = 1$.

Since there are no further cases this shows Step 2 (it remains to check the matrices in case 4 and 5 – see exercise sheet 1) and therefore part (c) of the theorem. □

Step 3. Let $z \in D$ such that the stabilizer G_z of z is not $\pm\mathrm{Id}$. Then $z = i$, $z = \rho$ or $z = \rho + 1$.

Proof of Step 3. This follows directly from Step 2 by checking $gz = z$ for all possible g's. Step 3 proves part (d) of the theorem. ☐

Step 4. It remains to show that $SL_2(\mathbb{Z})$ is generated by S and T.

Proof of Step 4. Let $g \in G$ and let z be an arbitrary point of the interior of D. Then $gz \in \mathcal{H}$ and by Step 1 exists $g' \in G'$ such that $g'(gz) \in D$. Moreover Step 2 implies that either $g'g \in \{\pm \mathrm{Id}\}$ or z is on the boundary of D which is by assumption not the case. Thus either $g'g = \mathrm{Id}$ or $g'g = -\mathrm{Id}$. Since $S^2 = -\mathrm{Id} \in G'$, we deduce that $g \in G'$, so $\mathrm{SL}_2(\mathbb{Z})$ is generated by S and T. This proves part (a) of the theorem. ☐

Therefore the theorem is proved. ☐

Remark 1.2.3. We have seen in the proof of Theorem 1.2.2 that $\mathrm{SL}_2(\mathbb{Z})$ is generated by the elements S and T. These satisfy the relations

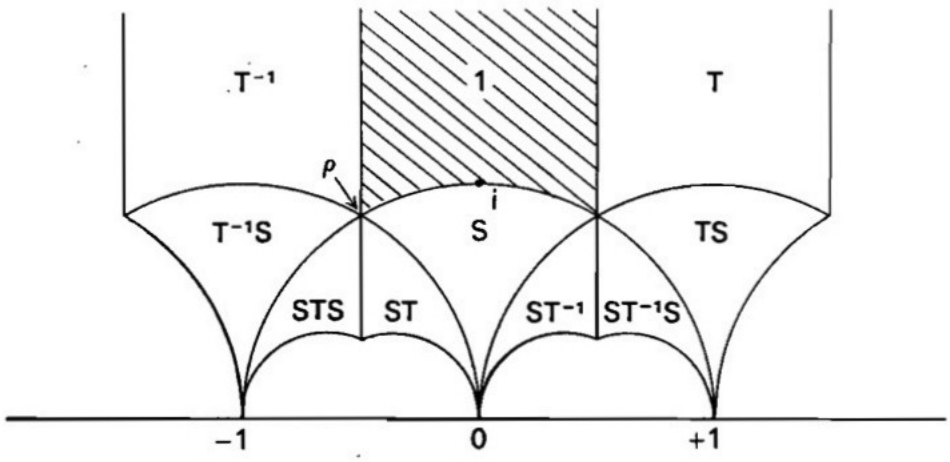
$$S^4 = \mathrm{Id} \quad (ST)^3 = S^2,$$

and one can show that these generate all the relations, i.e. that

$$\langle S, T \,|\, S^4,\, S^{-2}(ST)^3 \rangle$$

is a presentation of the group $\mathrm{SL}_2(\mathbb{Z})$.

Remark 1.2.4. The set D is called the **fundamental domain**. The figure below represents D itself and the transforms of D by some group elements of $\mathrm{SL}_2(\mathbb{Z})$. Part (c) of the theorem shows that two sets gD and $g'D$ where $g, g' \in \mathrm{SL}_2(\mathbb{Z})$ are either equal (if $g' = \pm g$) or only intersect along their edges. Furthermore part (a) implies that $\mathcal{H}$ is covered by the sets $\{gD \colon g \in \mathrm{SL}_2(\mathbb{Z})\}$: they form a **tesselation** of $\mathcal{H}$.



1.3 Modular forms and modular functions

Definition 1.3.1. A **weakly modular function of weight** k **and level 1** is a meromorphic function $\mathcal{H} \to \mathbb{C}$ such that $f|_k\alpha = f$ for all $\alpha \in \mathrm{SL}_2(\mathbb{Z})$, or equivalent

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k\ f(z)$$

for all $z \in \mathcal{H}$ and for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$.

Note 1.3.2. Since $\mathrm{SL}_2(\mathbb{Z})$ is generated by the matrices S and T, it is sufficient to check invariance under these two matrices, i.e. that

$$f(z+1) = f(z) \qquad \text{and} \qquad f(-1/z) = z^k f(z)$$

for all $z \in \mathcal{H}$.

Lemma 1.3.3. *There are no nonzero weakly modular functions of odd weight.*

Proof. Let k be odd and let f be a weakly modular function of weight k. As shown in (2) we have $f(z) = f(z+1)$ for all $z \in \mathcal{H}$. Moreover we get $f(z) = -f(z+1)$ for all $z \in \mathcal{H}$, since $f|_k\left(\begin{smallmatrix} -1 & -1 \\ 0 & -1 \end{smallmatrix}\right) = -f(\cdot + 1)$. So $f(z) = -f(z)$ and thus $f(z) = 0$ for all $z \in \mathcal{H}$. □

Define the function

$$\begin{aligned} q:\ \mathcal{H} &\to \mathbb{C}, \\ z &\mapsto \exp(2\pi i z). \end{aligned}$$

Note 1.3.4. Now let f be weakly periodic of weight k. Then f is periodic with period 1, so it can be written in the form

$$f(z) = \tilde{f}(\exp(2\pi i z)),$$

where $\tilde{f}$ is a meromorphic function on the punctured unit disk

$$\mathbb{D}^* = \{q \in \mathbb{C} :\ 0 < |q| < 1\}.$$

Note 1.3.5. The function $\tilde{f}$ is defined by

$$\tilde{f}(q) = f\left(\frac{\log q}{2\pi i}\right).$$

Observe that the logarithm is multi-valued, but choosing a different value of the logarithm is the same as adding an integer to $\frac{\log q}{2\pi i}$. The periodicity of f hence implies that $\tilde{f}(q)$ does not depend on the chosen value of the logarithm.

Note 1.3.6. Any weakly modular function can be written as

$$f(z) = \sum_{n=-\infty}^{\infty} a_n q^n$$

for some $a_n \in \mathbb{C}$ where $q = e^{2\pi i z}$; we call this the *q-expansion of f*. This is just the Laurent series of $\tilde{f}$ around $q = 0$, which converges for $0 < |q| < \varepsilon$ for ε sufficiently small ($\Leftrightarrow \Im(z) \gg 0$)

Definition 1.3.7.

- We say that f is meromorphic at ∞ if $a_n = 0$ for $n < -N$ and some $N \in \mathbb{N}$.
- We say that f is holomorphic at ∞ if $a_n = 0$ for $n < 0$. In this case, we define the value of f at ∞ to be $f(\infty) = \tilde{f}(0) = a_0$.

Definition 1.3.8. Let f be a weakly modular function of weight k and level 1.

1. If f is meromorphic on $\mathcal{H} \cup \{\infty\}$ we say f is a **modular function** (of weight k and level 1).
2. If f is holomorphic on $\mathcal{H} \cup \{\infty\}$ we say f is a **modular form** (of weight k and level 1).
3. If f is holomorphic on $\mathcal{H} \cup \{\infty\}$ and $f(\infty) = 0$ we say f is a **cuspidal modular form** or **cusp form**.

Note 1.3.9. If f and g are modular forms (resp. modular functions) of level 1 and weights k and ℓ, then the product fg is a modular form (resp. modular function) of weight $k + \ell$.

1.4 Eisenstein series

Definition 1.4.1. Let $k \geq 4$ even. Define the **Eisenstein series of weight** k to be the function $G_k \colon \mathcal{H} \to \mathbb{C}$ given by

$$G_k(z) = \sum_{(m,n)\in\mathbb{Z}^2\setminus\{0\}} \frac{1}{(mz+n)^k}. \tag{1.1}$$

Recall the following result from complex analysis:

Proposition 1.4.2. *Let U be an open subset of $\mathbb{C}$, and let $(f_n)_n \geq 0$ be a sequence of holomorphic functions on U that converges uniformly on compact subsets of U. Then the limit function $U \to \mathbb{C}$ is holomorphic.*

Lemma 1.4.3. *The series defining $G_k(z)$ converges absolutely and uniformly on subsets of $\mathcal{H}$ of the form*

$$R_{r,s} = \{x + iy : |x| \leq r, y \geq s\}.$$

It hence converges to a holomorphic function on $\mathcal{H}$.