FS 2025

Modular forms

1. Introduction

In simplest terms, modular forms (also called automorphic forms) are certain holomorphic functions which have "many symmetries", somewhat analogous to periodicity for functions defined on $\mathbb{R}$. But this definition (which will be given formally below) does not explain why these functions and their generalizations play such a central role in modern number theory, and why they are also linked in very surprising ways to many other topics in mathematics.

The goal of this course is to present the basic definitions and examples of modular forms, and to sketch some of their applications and interactions.

We begin with a list of results which do not (in any way) mention modular forms, or seem entirely far away even from complex analysis or sometimes number theory, and yet are examples of applications of modular forms (and often ~~the~~ cannot be proved, in the current state of knowledge, without them).

(1) Fermat's Great Theorem : if $n \geq 3$ then the equation $x^n + y^n = z^n$ has no integral solution with $xyz \neq 0$. This was proved by Wiles and Taylor-Wiles ; the connection with modular forms had been discovered previously by Frey, Hellegouarch, Ribet, and work of Serre also played a big part.

(2) Optimal sphere packings : in $\mathbb{R}^n$, one can try to arrange disjoint unit spheres so that their union covers as much space as possible ; finding the best way to do this is a big problem and for $n = 8$ and 24, M. Viazovska ~~[illegible]~~ proved that known constructions (related to special lattices) were optimal [~~—~~ jointly with Cohn, Kumar, Miller for $n = 24$]

(3) A problem of Banach and Ruziewicz

Question: is there a finitely-additive measure μ on the unit sphere $S_n \subset \mathbb{R}^{n+1}$ which is invariant by rotation, defined on Lebesgue-measurable sets, of total measure 1 and <u>not</u> the Lebesgue measure?

Here, "finitely-additive" means that

$$\mu\left(\bigcup_{i\in I} X_i\right) = \sum_{i\in I} \mu(X_i)$$

holds, for I <u>finite</u> (and X_i measurable). Banach (1923) proved that there <u>are</u> such measures on S^1. The other cases remained open until 1984, when Drinfeld showed no such measure exists if $n = 2, 3$ (the case $n \geq 4$ was already known from work of Margulis and Sullivan in 1980-81); Drinfeld uses modular forms to do this.

(4) Distribution of points on spheres

Another question about spheres goes back to Linnik:

Q. Given $k \geq 1$, define for $N \geq 1$

$$X_N = \{(n_1, \dots, n_k) \in \mathbb{Z}^k \mid n_i \text{ integer}, n_1^2 + \dots + n_k^2 = N\}$$

[integral points on the sphere of radius $\sqrt{N}$ in $\mathbb{R}^n$]. Are the points $X_N/\sqrt{N}$ equidistributed on the unit sphere? In other words, is it true that for $f: S^{n-1} \to \mathbb{C}$ continuous, we have

$$\lim_{N\to\infty} \frac{1}{|X_N|} \sum_{x\in X_N} f\left(\frac{x}{\sqrt{N}}\right) = \int f \, d\mu$$

where μ is the Lebesgue measure on the sphere?

Linnik proved that this is true for $n \geq 4$ and we will give a proof using modular forms. For $n=3$, this is also true but much more difficult; known proofs all use modular forms (the first is due to Iwaniec and Duke in 1987-88).

(5) Distribution of roots of quadratic congruences

Here we pick an integer $d \in \mathbb{Z}$, squarefree, and we consider the distribution of the roots of $x^2 - d = 0$ modulo p as p varies. From quadratic reciprocity and Dirichlet's Theorem on primes in arithmetic progressions, it follows that

{ There are two roots $x_p, -x_p$ modulo p for about half the primes
{ There are no roots for the other half

Q. How are $x_p, -x_p$ distributed if we think of them as integer in $\{0, \dots, p-1\}$? Do they get "bunched up" in some subinterval?

Theorem (Duke, Friedlander, Iwaniec, 1995 Tóth 1997)

For $f: [0,1] \to \mathbb{C}$ continuous, we have

$$\int_0^1 f(t)\,dt = \lim_{x\to\infty} \frac{1}{\pi(x)} \sum_{p\le x} \sum_{\substack{x^2-d\equiv 0 \,(p)\\ 0\le x<p}} f\left(\left\{\frac{x}{p}\right\}\right)$$

In other words, the roots are, after normalizing, equidistributed (w.r.t the Lebesgue measure).

(6) Construction of Ramanujan graphs (Lubotzky - Phillips - Sarnak; Margulis)

(7) The problem of congruent numbers (Tunnell)

(8) Number of zeros of $\zeta(s)$ with $\mathrm{Re}(s) = 1/2$, $|\mathrm{Im}(s)| \le T$ (Dehouillen - Iwaniec; Conrey)

2. ("Classical") modular forms

We begin with the basic definitions...

Definition (holomorphic modular form)

(1) $\mathbb{H} = \{ z \in \mathbb{C} \mid \mathrm{Im}(z) > 0 \}$
(open subset of $\mathbb{C}$, called the Poincaré upper-half plane)

(2) $SL_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{R}) \;\middle|\; ad - bc = 1 \right\}$

$GL_2^+(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{R}) \;\middle|\; ad - bc > 0 \right\}$

$SL_2(\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{R}) \;\middle|\; a, b, c, d \text{ in } \mathbb{Z} \right\}$

(3) A (holo/mero)morphic function

$$f : \mathbb{H} \longrightarrow \mathbb{C}$$

is a (level 1) modular form of weight k, where $k \in \mathbb{Z}$, if

$\forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$, $f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k f(z)$.
$\forall z \in \mathbb{H}$

We now need to spend some time to understand why this even makes sense, and what this really means.

(1) If $z \in \mathbb{H}$, $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{R})$, then $\frac{az+b}{cz+d} \in \mathbb{H}$: in fact, check that

$$\operatorname{Im}\left(\frac{az+b}{cz+d}\right) = \frac{\operatorname{Im}(z)}{|cz+d|^2}$$

(uses $ad-bc=1$).

(2) Even more precisely, the definition

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

defines a group action of $SL_2(\mathbb{R})$ on $\mathbb{H}$:

we have $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot z = z$ for all $z \in \mathbb{H}$

$$g_1 \cdot (g_2 \cdot z) = (g_1 g_2) \cdot z \quad \text{for all } g_i \in SL_2(\mathbb{R}),\ z \in \mathbb{H}$$

This action is transitive: if $z = x+iy$, $y>0$ then

$$z = y\left(i + \frac{x}{y}\right) = \begin{pmatrix} \sqrt{y} & \\ & \frac{1}{\sqrt{y}} \end{pmatrix} \begin{pmatrix} 1 & \frac{x}{y} \\ 0 & 1 \end{pmatrix} i;$$

so the orbit of i is all of $\mathbb{H}$. We see here already two special subgroups of $SL_2(\mathbb{R})$:

$$A = \left\{ \begin{pmatrix} y & \\ & \frac{1}{y} \end{pmatrix} \mid y>0 \right\} \quad (\text{"Cartan subgroup"})$$

$$U = \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \mid x \in \mathbb{R} \right\} \quad (\text{"unipotent subgroup"})$$

Note $A \cdot U = U \cdot A = \left\{ \begin{pmatrix} x & y \\ 0 & x^{-1} \end{pmatrix} \mid \begin{matrix} x>0 \\ y \in \mathbb{R} \end{matrix} \right\} = B,$

(called a Borel subgroup), and we see that B already acts transitively.

For any transitive action of a group G on a set $X \neq \emptyset$, there are bijections

$$f_{x_0} \begin{cases} G \longrightarrow X \\ H_{x_0} g \longmapsto g x_0 \end{cases}$$

defined for any $x_0 \in X$, where

$$H_{x_0} = \{ g \in G \mid g \cdot x_0 = x_0 \}$$

is the stabilizer of x_0.

For $SL_2(\mathbb{R})$ acting on $\mathbb{H}$, we take $x_0 = i$:

$$\frac{ai+b}{ci+d} = i$$

$$\Updownarrow$$

$$ai + b = -c + id$$

i.e. $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$; then the determinant is $a^2 + b^2$, so we must have $a^2 + b^2 = 1$. In other words

$$\mathrm{Stab}_i(SL_2(\mathbb{R})) = SO_2(\mathbb{R})$$

$$= \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \middle| \; \theta \in \mathbb{R} \right\}$$

Lemma (Iwasawa decomposition, or NAK decomposition)

We have $SL_2(\mathbb{R}) = NAK$, with $N = U$, $K = \mathrm{Stab}_i = SO_2(\mathbb{R})$

Proof. Take $g \in SL_2(\mathbb{R})$; put $z = g \cdot i$. By the above we can find $n \in U$, $a \in A$ s.t. $nai = z = gi$. Then $k = (na)^{-1} g \in \mathrm{Stab}_i$ and $g = nak \in NAK$.

□

(3) Let $\mathcal{F} = \{ f : \mathcal{H} \longrightarrow \mathbb{C} \}$ and let $k \in \mathbb{Z}$ be an integer. Then defining

$$(f|_k g)(z) = (cz+d)^{-k} f(z)$$

("weight k action") defines an action (on the right) of $SL_2(\mathbb{R})$ on $\mathcal{F}$:

$$(f|_k g_1)|_k g_2 = f|_k (g_1 g_2) \qquad (*)$$

(Exercise)

Then: (i) to say that f is modular of weight k means that

$$\forall g \in SL_2(\mathbb{Z}), \quad f|_k g = f$$

(ii) In particular, it is enough to prove this for g in a generating set of $SL_2(\mathbb{Z})$ to know it for all of $SL_2(\mathbb{Z})$ [this is due to the action property $(*)$].

(4) Why should modular forms exist? Even finding functions in $\mathcal{F}$ with less regularity (weight k)

is not obvious, at first sight. But note that the action property means that an f which is modular is characterized uniquely by the values it takes on orbits of points of $\mathcal{H}$ under the action of $SL_2(\mathbb{Z})$ [not of $SL_2(\mathbb{R})$].

Here the analogy is with 1-periodic functions

$$f: \mathbb{R} \longrightarrow \mathbb{R}$$

i.e. f s.t. $f(x+1) = f(x)$ for all $x \in \mathbb{R}$, which means that f is invariant for the translation action

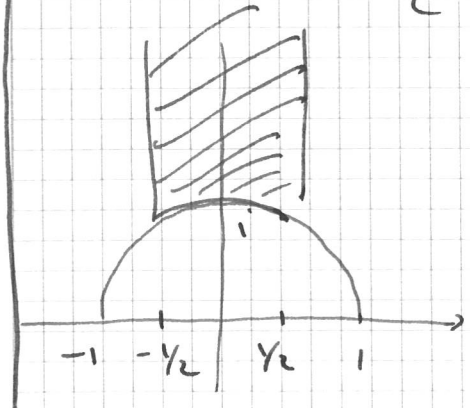
$$(t \cdot f)(x) = f(x+t)$$

of $\mathbb{Z} \subset \mathbb{R}$ on $\{f: \mathbb{R} \longrightarrow \mathbb{R}\}$. In this case we know that we can define f arbitrarily on $[0,1[$, and then "extend" by periodicity.

We can do something similar here.

Proposition. (Serre, Th. 1 in §VII.1.2)

Let $D = \left\{ z \in \mathcal{H} \;\middle|\; |z| \geq 1,\ |\mathrm{Re}(z)| \leq \frac{1}{2} \right\} \subset \mathcal{H}$.



Then (i) any element in $\mathcal{H}$ is of the form gz for some $g \in SL_2(\mathbb{Z})$, $z \in D$.

(ii) If the orbits of two points of D° intersect, then $z = z'$.

(iii) More generally, if z, z' in $\mathcal{D}$ are such that there is $g \in SL_2(\mathbb{Z})$ s.t. $gz = z'$ then

either $\mathrm{Re}(z) = -\frac{1}{2}$, $z' = z+1$

or $\mathrm{Re}(z) = \frac{1}{2}$, $z' = z-1$

or $|z| = 1$, $z' = -\frac{1}{z}$

(iv) For $z \in \mathcal{D}$, we have

$$\mathrm{Stab}_{SL_2(\mathbb{Z})}(z) = \{ g \in SL_2(\mathbb{Z}) \mid g \cdot z = z \}$$

$$= \begin{cases} \{\pm 1\}, & z \notin \{i, e^{2i\pi/3}, e^{\frac{2i\pi}{3}}\} \\ \pm \left\langle \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right\rangle, & z = i \\ \pm \left\langle \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} \right\rangle, & z = e^{2i\pi/3} \\ \pm \left\langle \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \right\rangle, & z = e^{i\pi/3} \end{cases}$$

Corollary - The elements

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

generate $PSL_2(\mathbb{Z})$; they satisfy the relations

$$S^2 = T^3 = \mathrm{Id}.$$

(Here $PSL_2(\mathbb{Z}) = SL_2(\mathbb{Z})/\{\pm 1\}$; note that

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \cdot z = \frac{-z}{-1} = z$$

for all $z \in \mathcal{H}$, so $-\mathrm{Id}$ acts trivially on $\mathcal{H}$, and hence the action of $SL_2(\mathbb{Z})$ defines by passing to the quotient an action of $PSL_2(\mathbb{Z})$ on $\mathcal{H}$.) Note also that $S \cdot z = -\frac{1}{z}$, $Tz = z+1$.

This Theorem means that a function defined on D, satisfying only the compatibility conditions "on the boundary" extends uniquely to a function on $\mathbb{H}$ satisfying

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

for $z \in \mathbb{H}$, $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$. (Getting a holomorphic or meromorphic function requires a bit more care.)

Proof. We will prove (i) and (ii) and the Corollary, leaving some details of (iii) and (iv) (see Serre's book).

The key point is the following fact: $SL_2(\mathbb{Z})$ is a discrete subgroup of $SL_2(\mathbb{R})$ [like $\mathbb{Z}$ is a discrete subgroup of $\mathbb{R}$].

Let Γ be the subgroup generated by $\pm Id$, S and T. We prove (i) with an element of Γ.

Claim: for $z \in \mathbb{H}$, there exists $g \in \Gamma$ such that $\mathrm{Im}(gz)$ is maximal.

Indeed, recall that

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

Now observe that there are only finitely many

$g \in \Gamma$ such that, writing $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, we have $|cz+d|^2 \leq 1$ [this is $(cx+d)^2 + c^2y^2$, so we must have $|c| \leq 1$ and $|cx+d| \leq 1$].

So there are only finitely many $g \in \Gamma$ with

$$\mathrm{Im}(gz) \geq \mathrm{Im}(z),$$

and we pick one giving the largest value of this set [which contains $g = Id$].

Since $\mathrm{Im}(Sgz) = \frac{\mathrm{Im}(gz)}{|gz|^2}$, we must have $|gz| \geq 1$. Now pick $k \in \mathbb{Z}$ such that

$$|\mathrm{Re}\, gz - k| \leq \frac{1}{2}$$

and note that

$$T^k g z = gz + k$$

to conclude that $T^k g z \in D$, with $T^k g \in \Gamma$. This proves (i) in stronger form (apparently).

Let now $\left| \begin{array}{l} z \in D \\ g \in SL_2(\mathbb{Z}) \end{array} \right.$ be such that $gz \in D$.

Suppose $\mathrm{Im}(gz) \geq \mathrm{Im}\, z$, i.e. $|cz+d| \leq 1$.

Then first we must have $|c| \leq 1$, as above, so either $c = 0$, and $g \cdot z = z + b$ is a translation, hence we get $\mathrm{Re}(z) = -\frac{1}{2}$

or $c = 1$ and $|cz+d| = |z+d| \leq 1$

implies $\begin{cases} d = 0 \\ \text{or} \\ z = e^{2i\pi/3}, \end{cases}$ $d = 1, \ldots$

If $\operatorname{Im} gz < \operatorname{Im} z$, we replace (z, g) by (gz, g^{-1}) [using $g.z \in D$].

Now using this we conclude that in fact

$$\Gamma = SL_2(\mathbb{Z})$$

Indeed, let $g \in SL_2(\mathbb{Z})$. Then $g.(2i) \in \mathbb{H}$ so by (i) [in strong form] we find $h \in \Gamma$ s.t. $h.g.(2i) \in D$. Since $2i \in D$, the property (ii) implies $hg = \pm Id$, so $g \in \Gamma$.

□