April 30, 2:00 pm

1 Second Facts About Spaces of Modular Forms

We have repeatedly used facts about the dimensions of the space of modular forms, mostly to give specific examples of and relations between modular forms of a given weight. We now prove these results in this section.

Recall that a modular form f of weight 0 is a holomorphic function that is invariant under the action of $SL(2,\mathbb{Z})$ and whose domain can be extended to include the point at infinity. Hence, we may regard f as a holomorphic function on the compact Riemann surface $G\backslash\mathcal{H}^*$, where $\mathcal{H}^* = \mathcal{H} \cup \{\infty\}$. If you haven't had a course in complex analysis, you shouldn't worry too much about this statement. We're really just saying that our fundamental domain can be understood as a compact manifold with complex structure. So locally, we look like $\mathbb{C}$ and there are compatibility conditions on overlapping local maps.

By the maximum modulus principle from complex analysis (see section 3.4 of Ahlfors for example), any such holomorphic function must be constant. We can use this fact, together with the fact that given $f_1, f_2 \in \mathcal{M}_{2k}$, then f_1/f_2 is a meromorphic function invariant under G (i.e. a weakly modular function of weight 0).

Usually, one proves facts about the dimensions of spaces of modular forms using foundational algebraic geometry (the Riemann-Roch theorem, in particular) or the Selberg trace formula (an even more technical result relating differential geometry to harmonic analysis in a beautiful way). These apply to spaces of modular forms for other groups, but as we're only going to say a few words about such spaces, we won't delve into either, but rather prove a much simpler result in the spirit of the Riemann-Roch theorem.

Proposition 1 *Let X be a compact Riemann surface. Fix a collection of points $P_1, \ldots, P_n \in X$ and positive integers $r_1, \ldots, r_n$. Define V to be the vector space of meromorphic functions on X which are holomorphic except possibly at the P_i, and at each P_i, are either holomorphic or have a pole of order at most r_i. Then*

$$\dim(V) \leq r_1 + r_2 + \cdots + r_n + 1$$

Proof For each j with $1 \leq j \leq n$, pick coordinate functions $t = t_j$ in a neighborhood of P_j so that (with respect to t_j) P_j is the origin. (This is precisely what we're doing for modular forms when we take $q = e^{2\pi i z}$. This coordinate q takes the point at infinity $i\infty$ to the origin.) Then any meromorphic function ϕ in V has Laurent expansion in $t = t_j$ of form

$$\phi(t) = a_{-r_j} t^{-r_j} + a_{-r_j+1} t^{-r_j+1} + \cdots$$

To each ϕ, we collect all the negative coefficients from the Laurent expansions for each coordinate $t = t_j$. There are $r := r_1 + \cdots + r_n$ of these, which we record as a vector $v(\phi) \in \mathbb{C}^r$. Given $\phi_1, \ldots, \phi_N$ in V with $N > r$, then there exist coefficients $c_1, \ldots, c_N \in \mathbb{C}$, not all 0, for which

$$c_1 v(\phi_1) + c_2 v(\phi_2) + \cdots + c_N v(\phi_N) = 0.$$

But then $c_1\phi_1 + \cdots + c_N\phi_N$ has no poles (as we've removed the poles at all possible P_j). Again, applying the maximum modulus principle, such a function must be constant. So any subspace of V with dimension $> r$ contains a constant function, hence $\dim(V) \leq r + 1$. □

Proposition 2 *The space $\mathcal{M}_{2k}$ is finite dimensional.*

Proof Given a non-zero element g of $\mathcal{M}_{2k}$, let X be the compact Riemann surface formed by $G\backslash\mathcal{H}^*$ and let $P_1, \ldots, P_n$ be the zeros of g, with orders $r_1, \ldots, r_n$. (Technicality: Some care needs to be taken in how we count these orders, as some points have non-trivial stabilizers, like $i, \rho = e^{2\pi i/3}, \bar{\rho}$. If there are zeros at these points, their orders should be multiplied by the size of the stabilizer (which are 2 or 3, resp.).) Let V be the vector space as in the previous proposition formed with the data P_j and r_j's. If $f \in \mathcal{M}_{2k}$, then f/g is a weakly modular function of weight 0. In particular, the map $f \mapsto f/g$ is an isomorphism of $\mathcal{M}_{2k}$ with V and hence has dimension at most $r + 1$. □

We now prove a much more precise result about the dimension of spaces of modular forms.

Theorem 1 *Write $2k = 12j + r$ with $0 \leq r \leq 10$. Then*

$$\dim(\mathcal{M}_{2k}) = \begin{cases} j+1 & \text{if } r = 0, 4, 6, 8, 10 \\ j & \text{if } r = 2 \end{cases}$$

More succinctly, the ring $\bigoplus_{k=0}^{\infty} \mathcal{M}_{2k}$ is generated by G_2 and G_3, the Eisenstein series of weights 4 and 6, respectively.

Proof First we show $\dim(\mathcal{M}_{2k}) = 1$ for $2k = 4, 6, 8$ or 10. For such k, suppose f is not in the one-dimensional space generated by G_k. By subtracting an appropriate multiple of G_k, we may assume that f has constant term 0 in its q-expansion (i.e. is a cusp form). Consider the function $G_h(f/\Delta)^6$, where $2h = 6(12 - 2k)$ and Δ is the cusp form of weight 12 given by the Ramanujan discriminant function. This is a modular form of weight 0 with no poles (the product form of Δ shows that it is never 0 except at ∞, where f too is 0) and hence is constant. Thus $G_h = c(\Delta/f)^6$ for some constant c, and hence G_h can have no zeros on $\mathcal{H}$. But then Δ^m/G_h with $m = 6h$ is a modular form of weight 0 with no poles and a zero of order m at ∞, a contradiction (the zeros and poles must be equal, counting with multiplicity). Hence, $\dim(\mathcal{M}_{2k}) = 1$ for $2k = 4, 6, 8$ or 10.

To show $\mathcal{M}_2$ is 0, suppose f is a non-zero modular form of weight 2. Then from what we've just shown, $fG_2 = cG_3$ for some constant c. But then since $G_2(e^{2\pi i/3}) = 0$, we would have $G_3(e^{2\pi i/3}) = 0$ and hence $\Delta(e^{2\pi i/3}) = 0$ since Δ is a linear combination of G_2^3 and G_3^2, a contradiction since Δ is never 0. So we're done for $k < 12$ (noting that $\dim(\mathcal{M}_0) = 1$, containing the constants).

For $k \geq 12$, we recall that $\dim(\mathcal{S}_{2k}) = \dim(\mathcal{M}_{2k}) - 1$. Moreover, multiplication by Δ is an isomorphism between $\mathcal{M}_{2k-12}$ and $\mathcal{S}_{2k}$. (It's clearly injective, and if $f \in \mathcal{S}_{2k}$, then f/Δ again has no poles, so is in $\mathcal{M}_{2k-12}$.

To show that G_2 and G_3 generate, first note that $\mathcal{M}_8$ and $\mathcal{M}_{10}$ are one dimensional, so G_4 and G_5 must be linear combinations of G_2^2 and G_2G_3, respectively. This finishes the proof for $k \leq 10$. But Δ is generated by G_2 and G_3 as noted above, so bootstrapping from $k \leq 10$, we have $\Delta\mathcal{M}_{2k-12} = \mathcal{S}_{2k}$ generated by them as well. Finally, noting that $G_2^r G_3^s$ with $4r + 6s = 2k$ is not cuspidal and in $\mathcal{M}_{2k}$ finishes the claim. □

2 Modular Forms for Other Groups

We've been dealing with modular forms for the group $SL(2, \mathbb{Z})$, or more properly the modular group G. This group arose naturally for us, relating to the equivalence on basis vectors for lattices. But suppose we want to generalize this notion of modular forms to groups other than G. We'll need to take stock of which properties of G were necessary in formulating reasonable definitions for modular forms.

Recall that the action of matrices γ on points in the upper half plane is well-defined for any $\gamma \in SL(2, \mathbb{R})$. In fact $SL(2, \mathbb{R})$ acts transitively on $\mathcal{H}$ because the group of upper triangular matrices B acts transitively (i.e. takes any element in $\mathcal{H}$

to any other element by an upper-triangular matrix):

$$\begin{pmatrix} y^{1/2} & xy^{-1/2} \\ 0 & y^{-1/2} \end{pmatrix} : i \mapsto x + iy$$

so every element in $\mathcal{H}$ is in the orbit of i. The stabilizer of i is (check this for yourself)

$$SO(2) = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \mid a^2 + b^2 = 1 \right\}$$

whose entries are often represented by $\sin\theta$ and $\cos\theta$ for a single parameter θ. So $\mathcal{H}$ may be identified with cosets $SL(2,\mathbb{R})/SO(2)$. Further, since B acts transitively on $\mathcal{H}$, it can be made to act transitively on $SL(2,\mathbb{R})/SO(2)$. Equivalently, we may write $SL(2,\mathbb{R}) = B{\cdot}SO(2)$; this is known as the Iwasawa decomposition for $SL(2,\mathbb{R})$. This is the precise sense in which we mean that $\mathcal{H}$ is a model for $SL(2,\mathbb{R})$.

To define modular forms, we considered a particular discrete subgroup $SL(2,\mathbb{Z})$ in $SL(2,\mathbb{R})$. We then identified $SL(2,\mathbb{Z})\backslash\mathcal{H}$ with a fundamental domain D whose boundary was formed by the intersection of geodesics in the hyperbolic plane. (Remember, these geodesics for $\mathcal{H}$ are either vertical lines starting at the real line in $\mathbb{C}$, or they're semicircles intersecting $\mathbb{R}$ at right angles. This comes from computing the shortest distance between two points with the metric $dx\,dy/y^2$ where $z = x + iy$.) Modular forms $f(z)$ for $SL(2,\mathbb{Z})$ of weight $2k$ could then be understood as invariant differential forms $f(z)dz^k$ on the compact Riemann surface built from the fundamental domain D.

In choosing a different subgroup of $SL(2,\mathbb{R})$, we'd like to retain the same properties, most importantly having a fundamental domain realized as a hyperbolic polygon (i.e. bounded by geodesics. In Euclidean geometry, geodesics are of course just straight lines, so "hyperbolic polygon" generalizes the notion of usual polygon.) In a moment, we'll give some explicit examples and deal almost exclusively with them, but first a word about the general problem. On a first reading, one may initially want to skip these generalities.

The group $SL(2,\mathbb{R})$ can be embedded in $M_2(\mathbb{R})$, the group of all 2×2 real matrices. This has an inner product $\langle g, h\rangle = \mathrm{Trace}(gh^t)$ for $g, h \in M_2(\mathbb{R})$ so that

$$\langle g, g\rangle = ||g||^2 = a^2 + b^2 + c^2 + d^2 \quad \text{for } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

This defines a topology on $M_2(\mathbb{R})$ and $SL(2,\mathbb{R})$ inherits a topology induced from the embedding.

Definition 1 *A subgroup* $\Gamma \subset SL(2,\mathbb{R})$ *is "discrete" if the induced topology is discrete, i.e.*

$$\{\gamma \in \Gamma \mid ||\gamma|| < C\} \text{ is a finite set for any constant } C > 0.$$

Definition 2 *Let* X *be a topological space with an action of a group* Γ *giving homeomorphisms of* X. *Then* Γ *"acts discontinuously on* X*" if the orbit* $\Gamma \cdot x$ *of any* $x \in X$ *has no limit point in* X.

The following old result of Poincaré connects these two definitions in our case.

Proposition 3 (Poincaré) *A subgroup of* $SL(2,\mathbb{R})$ *is discrete if and only if, when considered as a subgroup of* $PSL(2,\mathbb{R}) = SL(2,\mathbb{R})/\{\pm 1\}$, *it acts discontinuously on* $\mathcal{H}$.

Such a subgroup of $PSL(2,\mathbb{R})$ is called a Fuchsian group. Any Fuchsian group Γ can be shown to have a fundamental domain D. Recall, a fundamental domain for Γ has three properties:

- D is a domain (connected, open set) in $\mathcal{H}$.
- Distinct points of D are not equivalent under Γ.
- The orbit of any point $z \in \mathcal{H}$ contains a point in the closure of D.

These domains D are not unique, but one can show they all have the same volume (possibly infinite). We'd like the volume to be finite so that we can integrate over D AND we'd like to be able to choose D to be a (hyperbolic) polygon. For this, we need a final assumption:

Definition 3 *A Fuchsian group* Γ *is "of the first kind" if every point on the boundary of* $\mathcal{H}$, $\partial\mathcal{H} = \mathbb{R} \cup \{\infty\}$, *is a limit point of the orbit* $\Gamma \cdot z$ *for some* $z \in \mathcal{H}$.

One can show that $SL(2,\mathbb{Z})$ is a Fuchsian group of the first kind (see Bump for details). Moreover, any finite index subgroup of a Fuchsian group of the first kind is again such a group. There are many nice results for such groups, but here's just one (whose proof is due to Siegel):

Theorem 2 *Every Fuchsian group of the first kind* Γ *has a fundamental domain* D *which is a hyperbolic polygon with an even number of sides. The sides of* D *can be arranged in pairs, whose points are equivalent under* Γ *and the elements of* Γ *that pair each side generate* Γ *(hence a finite number of generators). Moreover,* D *has finite volume.*

3 Modular Forms for Congruence Subgroups

As we laid out in sketch in the previous section, Fuchsian groups of the first kind have all the right properties for defining modular forms. We briefly noted there that $SL(2,\mathbb{Z})$ and its subgroups of finite index are Fuchsian groups of the first kind. There is a nice family of groups of finite index, defined by

$$\Gamma(n) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \mid \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{n} \right\}.$$

Sometimes we write $SL(2,\mathbb{Z}) = \Gamma(1)$. Any subgroup of finite index Γ with

$$\Gamma(n) \subseteq \Gamma \subseteq \Gamma(1) = SL(2,\mathbb{Z}) \text{ for some } n$$

is called a congruence subgroup of $SL(2,\mathbb{Z})$. If Γ has index m, then one can make a fundamental domain D_Γ from the fundamental domain D for $SL(2,\mathbb{Z})$ by

$$D_\Gamma = \bigcup_{j=1}^{m} \gamma_j(D) \text{ for certain } \gamma_j \in \Gamma(1).$$

One should pick the γ_j to be distinct coset representatives for Γ in $\Gamma(1)$ for which D_Γ is a domain. There is a java applet online (very easy to play around with, and great for getting intuition about the shape of these fundamental domains) at

`http://www.math.lsu.edu/~verrill/`

for drawing these fundamental domains for $\Gamma(n)$ as well as the congruence subgroups

$$\Gamma_0(n) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \mid c \equiv 0 \pmod{n} \right\}.$$

Of particular importance are the images of $\{\infty\}$ under γ_j. Recall that

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : z \mapsto \frac{az+b}{cz+d} \quad \text{and} \quad \gamma(\infty) = \frac{a}{c} \in \mathbb{Q}.$$

This will give a set of inequivalent rational numbers $p_1, \ldots, p_r$ called cusps (where we take $p_1 = \infty$). We are now ready to define modular forms for a congruence subgroup.

Definition 4 *A modular form of weight k for Γ, a congruence subgroup of $\Gamma(1)$, is a holomorphic function $f(z)$ on $\mathcal{H}$ satisfying*

1. *$f(\gamma z) = (cz+d)^k f(z)$ for all $\gamma \in \Gamma$*

2. *$f(z)$ is holomorphic at EACH cusp p_i.*

This second item requires a bit more explanation. Just as $q = e^{2\pi i z}$ mapped the point ∞ to the origin in the coordinate q for $SL(2,\mathbb{Z})$, for our subgroup Γ, there are coordinate functions which take each cusp p_i to the origin, and for which we again have a Fourier expansion. This is usually achieved by mapping each p_i to ∞ by a matrix γ_i and then composing with $q = e^{2\pi i z/M}$ where

$$\gamma_i \Gamma \gamma_i^{-1} \cap \left\langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} 1 & M \\ 0 & 1 \end{pmatrix} \right\rangle$$

(as we can no longer guarantee that the function is periodic with period 1, since the matrix T may not be in Γ, but the fact that Γ is of finite index guarantees that some translation by an integer M will be a member of $\gamma_i \Gamma \gamma_i^{-1}$, and hence any modular form f for Γ will satisfy $f(z+M) = f(z)$ and be expressible in a Fourier expansion.)

4 Half-integral weight modular forms

Recall that we previously defined (for the lattice $\mathbb{Z} \subset \mathbb{R}$)

$$\theta(t) = \sum_{m=-\infty}^{\infty} e^{-\pi m^2 t} \quad t > 0$$

which we can extend to a definition for any $z \in \mathcal{H}$ by

$$\theta(z) = \sum_{m=-\infty}^{\infty} e^{i\pi m^2 z} \quad z \in \mathcal{H},$$

(which will still define an absolutely convergent function, according to the arguments we presented earlier for the convergence of the theta function.) Then we have the following transformations of $\theta(z)$:

$$\theta(z+2) = \theta(z), \quad \theta(-1/z) = \sqrt{-iz}\,\theta(z)$$

where the first equality is clear from the definition, and the second follows from Poisson summation along the imaginary axis, and analytic continuation. Let us choose the usual branch for the square root function, so that it is positive on positive reals. While these transformation laws implicitly suggest how θ transforms under any matrix in the group generated by $\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, they don't give an explicit transformation in terms of an arbitrary matrix. We want such an explicit transformation in order to suggest the appropriate definition of a half-integral weight modular form.

To this end, consider $\begin{pmatrix} b & -a \\ d & -c \end{pmatrix} \in SL(2, \mathbb{Z})$ with $b \equiv 0 \ (2), c \equiv 0 \ (2)$ (this somewhat strange choice will be justified later in the computation). Further take $d > 0$ (since $d < 0$ follows similarly). Then

$$\begin{aligned} \theta\left(\frac{bz-a}{dz-c}\right) &= \theta\left(\frac{b}{d} - \frac{1}{d(dz-c)}\right) \\ &= \sum_{m \ (\mathrm{mod}\ d)} e^{i\pi m^2 b/d} \sum_{t=-\infty}^{\infty} e^{-i\pi(\frac{m}{d}+t)^2 \frac{d}{dz-c}} \\ &= (id)^{-1/2}(dz-c)^{1/2} \sum_{m \ (\mathrm{mod}\ d)} e^{i\pi m^2 b/d} \sum_{u=-\infty}^{\infty} e^{2\pi i \frac{mu}{d} + iu^2\pi(z-\frac{c}{d})} \end{aligned}$$

where the last equality is obtained by applying Poisson summation to the inner sum in the second line. Interchanging the orders of summation, the inner sum over m (mod d) becomes

$$\sum_{m \ (\mathrm{mod}\ d)} e^{i\pi m^2 b/d + 2\pi i \frac{mu}{d}} = \sum_{m \ (\mathrm{mod}\ d)} e\left(\frac{\alpha m^2 + mu}{d}\right) \text{ with } e(x) = e^{2\pi i x} \text{ and } b = 2\beta, \beta \in \mathbb{Z}.$$

Since $-bc = -(2\beta)c \equiv 1 \pmod d$ and d odd (as $ad - bc = 1$) we may write $m = r\bar{\beta}$ mod d where $\beta\bar{\beta} \equiv 1 \ (d)$. Then

$$\begin{aligned} \sum_{m \ (\mathrm{mod}\ d)} e\left(\frac{\beta m^2 + mu}{d}\right) &= \sum_{r \ (\mathrm{mod}\ d)} e\left(\frac{\bar{\beta}(r^2 + ru)}{d}\right) \\ &= \sum_{r \ (\mathrm{mod}\ d)} e\left(\frac{\bar{\beta}(r + \bar{2}u)^2 - \bar{\beta}\bar{4}u^2}{d}\right) \\ &= e\left(\frac{-\bar{\beta}\bar{4}u^2}{d}\right) \sum_{r \ (\mathrm{mod}\ d)} e\left(\frac{\bar{\beta}r^2}{d}\right) \end{aligned}$$

where we've continued to write $\bar{x}$ for the inverse of x mod d. Substituting back into the sum over u above, we have

$$\theta\left(\frac{bz-a}{dz-c}\right) = (id)^{-1/2}(dz-c)^{1/2}\left[\sum_{r \pmod{d}} e\left(\frac{\bar{\beta}r^2}{d}\right)\right]\theta(z).$$

Substituting $-1/z$ for z in our equation for θ, we have

$$\begin{aligned}\theta\left(\frac{az+b}{cz+d}\right) &= (id)^{-1/2}(d(-1/z)-c)^{1/2}\left[\sum_{r \pmod{d}} e\left(\frac{\bar{\beta}r^2}{d}\right)\right]\theta(-1/z)\\ &= i(cz+d)^{1/2}d^{-1/2}\left[\sum_{r \pmod{d}} e\left(\frac{\bar{\beta}r^2}{d}\right)\right]\theta(z)\end{aligned}$$

where we used the transformation $\theta(-1/z) = \theta(z)\sqrt{-iz}$ in the last step.

Finally, recall that the sum in brackets is closely related to the Gauss sum

$$\sum_{m=0}^{d-1} e\left(\frac{m^2}{d}\right) = \begin{cases} d^{1/2} & \text{if } d \equiv 1 \pmod{4} \\ id^{1/2} & \text{if } d \equiv 3 \pmod{4}. \end{cases}$$

This evaluation is worked out carefully in many number theory textbooks, including Davenport's "Multiplicative Number Theory." Using this evaluation in the above, we have

$$\theta(\gamma(z)) = \left(\frac{2c}{d}\right)_2 \varepsilon_d^{-1}(cz+d)^{1/2}\theta(z)$$

for any $\gamma \in SL(2,\mathbb{Z})$ with $b, c \equiv 0\ (2)$, $\varepsilon_d = 1$ or i according to $d \equiv 1$ or 3 mod 4, respectively, and $\left(\frac{2c}{d}\right)_2$ is the Legendre symbol appropriately extended to include $|d| = 1$ (there are slight annoyances in this case, like $\left(\frac{0}{d}\right)_2 = 1$.)

Setting $\tilde{\theta}(z) = \theta(2z)$ we have just shown:

Theorem 3 *For any* $\gamma \in \Gamma_0(4)$*,*

$$\tilde{\theta}(\gamma(z)) = j(\gamma, z)\tilde{\theta}(z)$$

where

$$j(\gamma, z) = \left(\frac{c}{d}\right)_2 \varepsilon_d^{-1}(cz+d)^{1/2}$$

as defined above.

Any such function that transforms by $j(\gamma, z)^{2k}$ for k a fixed half-integer, and all $\gamma \in \Gamma_0(N)$ and is holomorphic at each cusp, is called a half-integral weight modular form. Note this definition is consistent with our earlier definitions for integral weight modular forms – if k is a proper even integer, $j(\gamma, z)$ reduces to the usual definition.

References

[1] D. W. Bump, *Automorphic Forms and Representations*, Cambridge Studies in Advanced Mathematics, vol. 55, Cambridge Univ. Press (1996).

[2] J.-P. Serre, *A Course in Arithmetic*, Graduate Texts in Mathematics, vol. 7, Springer-Verlag (1973).