MODULAR FORMS LECTURE 12: SPACES OF MODULAR FORMS

LARRY ROLEN, VANDERBILT UNIVERSITY, FALL 2020

> Without doubt it would be desirable to have a rigorous proof of this proposition; however I have left this research aside for the time being after some quick unsuccessful attempts, because it appears to be unnecessary for the immediate goal of my study.
>
> Riemann on his Hypothesis

1. Dimension Formulas

Just like we did with elliptic functions, modular forms will be determined by their **divisors**, the set of their poles and zeros. We can use this to study the spaces M_k and S_k, and in particular to give explicit formulas for their dimensions. To start, we can count the number of points in the divisor as a simple function of the weight.

Theorem (Valence Formula)**.** *Let $0 \neq f \in M_k^{mero}$. Then we have*

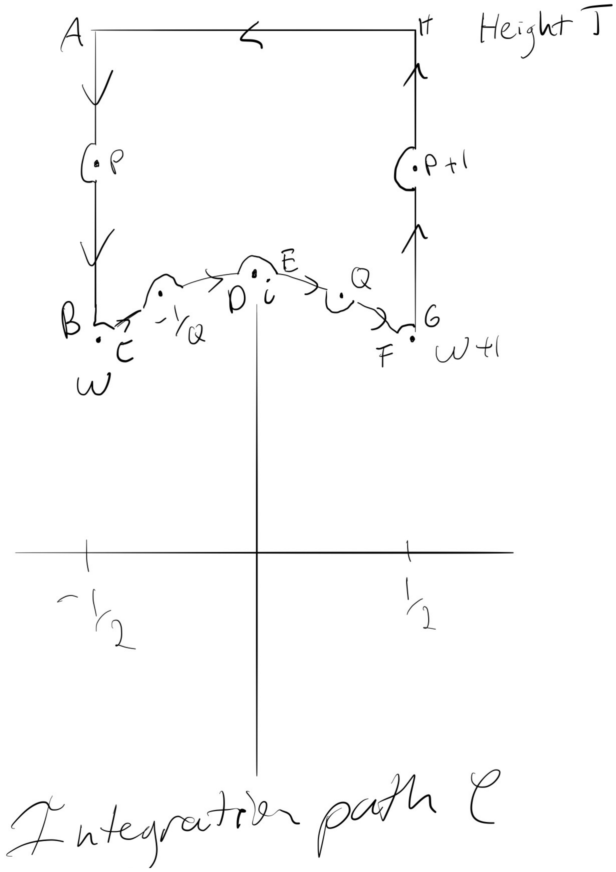
$$\nu_\infty(f) + \frac{1}{2}\nu_i(f) + \frac{1}{3}\nu_\omega(f) + \sum_{\tau_0 \in \mathcal{F}\setminus\{i,\omega\}} \nu_{\tau_0}(f) = \frac{k}{12}.$$

Here $\nu_\infty(f)$ is the **order of vanishing at ∞ of** *f (if $f(\tau) = \sum_{n \geq n_0} a_n q^n$ with $a_{n_0} \neq 0$, this is $\nu_\infty(f) = n_0$), and $\nu_{\tau_0}(f)$ is the order of vanishing at a point $\tau_0 \in \mathbb{H}$ (the Laurent expansion of f at τ_0 is of the form $\sum_{n \geq \nu_{\tau_0}(f)} b_n(\tau - \tau_0)^n$ with $b_{\nu(\tau_0)} \neq 0$. Note that the order is negative if there is a pole at a point.*

Proof. We want to count zeros and poles. As we did for elliptic functions, the go-to way to do this is via the **Argument Principle**; integrating the log derivative. We have to count zeros and poles inside of the fundamental domain, so we integrate along its boundary. Recall that the log derivative transforms zeros and poles of f into simple poles with residue equal to the order of vanishing.

We have to be careful with our path of integration, however. We need to avoid any points in the divisor of $\mathcal{F}$ which lie on the boundary. We also need a closed path, so we chop off the fundamental domain at some height T which is higher than any zeros or poles of f except a possible one at $i\infty$. This is possible, because of the condition that the function is meromorphic at $i\infty$. Specifically, as a function of q, this means that f is a meromorphic function on the punctured unit disk with the origin removed (this is one reason why some condition at $i\infty$ helps to obtain a good space of functions).

In the following picture, we draw a typical integration path that avoids a zero/pole on the side P (and the necessary additional zero/pole $TP = P + 1$), and a point on the bottom arc Q together with the point $SQ = -1/Q$. The small circular arcs are chosen so that one divisor point on the boundary lies inside of the curve and one lies on the outside, except that if there is a zero/pole at i or ω, then as in the picture they are left outside of the curve in all little circular arcs.



In short, the above picture shows the path when the zeros/poles on the boundary are P, Q, i, ω, and all zeros/poles on the interior have imaginary part less than T. The little arcs are circles of radius $\varepsilon > 0$, which will be taken to tend to zero.

The Residue Theorem/Argument Principle imply

$$\frac{1}{2\pi i}\int_{\mathcal{C}} \frac{f'(\tau)d\tau}{f(\tau)} = \sum_{\tau_0 \in \mathcal{F}\setminus\{i,\omega\}} \nu_{\tau_0}(f).$$

We will evaluate this integral piece by piece. Firstly, note that $f(\tau)$ being translation invariant implies that the logarithmic derivative is too, by differentiating the relation $f(\tau+1) = f(\tau)$ and using the (basically trivial here) chain rule. Thus the left and right pieces of the integral cancel out:

$$\int_{AB} \frac{f'(\tau)d\tau}{f(\tau)} + \int_{GH} \frac{f'(\tau)d\tau}{f(\tau)} = \int_{AB} \frac{f'(\tau)d\tau}{f(\tau)} - \int_{HG} \frac{f'(\tau)d\tau}{f(\tau)} = 0.$$

Next we look at the piece at the top, along HA. This is determined by the behavior at $i\infty$. Specifically, the line HA in the τ variable becomes the circle C_T of radius $e^{-2\pi T}$ centered at the origin in the q variable, **traversed clockwise**. Renaming this change of variables $\widetilde{f}(q) = f(\tau) = \sum_n a_n q^n$, and using

$$f'(\tau) = \frac{d}{dq}\widetilde{f}(q)\frac{dq}{d\tau}$$

we thus have

$$\frac{1}{2\pi i}\int_{HA}\frac{f'(\tau)d\tau}{f(\tau)} = \frac{1}{2\pi i}\int_{C_T}\frac{\widetilde{f}'(q)dq}{\widetilde{f}(q)}.$$

By the Residue Theorem, this is $-\nu_\infty(f)$ (the minus is because the path is clockwise).

We now evaluate the integrals by the points i and ω, namely the paths BC, DE, and FG. We will need a modification of the Residue Theorem. To explain this, we'll quickly recall how the Argument Principle works. If you have a function $f(\tau)$ with Laurent expansion $f(\tau) = c_m(z-a)^m + \dots$ with $c_m \neq 0$, then the Argument Principle gives us

$$\frac{f'(\tau)}{f(\tau)} = \frac{m}{z-a} + g(z),$$

where g is holomorphic at a. If you integrate this log derivative around a small circle of radius ε, going counter-clockwise, the integral will give you $2\pi i m$. If you instead only integrate over an arc of angle θ and let $\varepsilon \searrow 0$, then instead you get $\theta i m$. To see this, consider the basic example of the Residue Theorem (I previously mentioned this as an example to try explicitly if you want to see how it works). The integral of g on the tiny arc goes to zero, as the arc lengths do and it doesn't blow up near a. Thus, we want, after translating to get an integral of $1/z$ around the origin and parameterizing the circular arc as εe^{it} with $t \in [a, a+\theta]$

$$\lim_{\varepsilon\to 0}\int_a^{a+\theta}\frac{mi\varepsilon e^{it}dt}{\varepsilon e^{it}} = mi\lim_{\varepsilon\to 0}\int_a^{a+\theta}dt = mi\theta,$$

as claimed

In our situation, the integral along BC, with ε tending to 0, the limiting angle (between the line and circle pieces of the boundary) $\pi/3$ and the path is clockwise, and so

$$\int_{BC}\frac{f'(\tau)d\tau}{f(\tau)} = -\frac{1}{2\pi i}(\nu_\omega(f)\cdot \pi i/3) = -\frac{\nu_\omega(f)}{6}.$$

The integral along FG, in the same way, becomes $-\nu_{\omega+1}(f)/6 = -\frac{\nu_\omega(f)}{6}$. Similarly, the integral along DE becomes $-\frac{\nu_i(f)}{2}$, since the arc has angle tending to π.

We are now very close. We just have to show that as $\varepsilon \searrow 0$, we have

$$\frac{1}{2\pi i}\int_{CD}\frac{f'(\tau)d\tau}{f(\tau)} + \frac{1}{2\pi i}\int_{EF}\frac{f'(\tau)d\tau}{f(\tau)} = \frac{k}{12}.$$

The point is that $CD = -S\cdot(EF)$, that is, they map to each other under S with opposite orientation, but unlike with translation invariance, the modular inversion relation is

not preserved under differentiation. However, it nearly is. Let's look at what differentiation does to modularity. If

$$f(\gamma\tau) = (c\tau + d)^k f(\tau),$$

then we get

$$f'(\gamma\tau)\frac{d\gamma\tau}{d\tau} = (c\tau + d)^k f'(\tau) + kc(c\tau + d)^{k-1} f(\tau).$$

Dividing gives

$$\frac{f'(\gamma\tau)}{f(\gamma\tau)} d\gamma\tau = \frac{f'(\tau)}{f(\tau)} d\tau + k\frac{cd\tau}{c\tau + d}.$$

Thus, taking $c = 1$, $d = 0$, we get

$$\frac{1}{2\pi i}\int_{CD} \frac{f'(\tau)d\tau}{f(\tau)} + \frac{1}{2\pi i}\int_{EF} \frac{f'(\tau)d\tau}{f(\tau)} = \frac{1}{2\pi i}\int_{CD} \frac{f'(\tau)d\tau}{f(\tau)} - \frac{1}{2\pi i}\int_{S\cdot(CD)} \frac{f'(\tau)d\tau}{f(\tau)}$$

$$= \frac{1}{2\pi i}\int_{CD} \frac{f'(\tau)d\tau}{f(\tau)} - \frac{f'(S\tau)dS\tau}{f(S\tau)} = -\frac{k}{2\pi i}\int_{CD} \frac{d\tau}{\tau}.$$

As $\varepsilon \searrow 0$, this becomes (taking $z = e^{i\theta}$)

$$-\frac{k}{2\pi i}\lim_{\varepsilon\to 0}\int_{CD} \frac{d\tau}{\tau} = -\frac{k}{2\pi i}\lim_{\varepsilon\to 0}\int_{\frac{2\pi}{3}}^{\frac{\pi}{2}} \frac{ie^{i\theta}d\theta}{e^{i\theta}} = -\frac{k}{2\pi i}\cdot\left(-\frac{\pi i}{6}\right) = \frac{k}{12},$$

completing the proof.

□

Remark. *This is also a special case of a very general result called the* **Riemann-Roch theorem**. *When we look at more general situations, we may cite that, but it is worth doing this "hands-on" analysis proof once.*

Remark. *The proof suggests that derivatives of modular forms aren't quite modular forms, but nearly are. They are actually* **quasimodular forms**, *as we'll see.*

This allows us to prove very strong results on modular forms spaces. Before stating this, note that it is easy to check that $M_k M_\ell \subseteq M_{k+\ell}$; that is, a product of modular forms is a modular form where the weights add (this simply comes from the identity $(c\tau + d)^k(c\tau + d)^\ell = (c\tau + d)^{k+\ell}$).

Theorem. *Let k be an even integer. Then the following hold.*

1. $M_0 = \mathbb{C}$. *That is, holomorphic modular forms of weight 0 are constant.*
2. $M_k = 0$ *if* $k < 0$ *or* $k = 2$.
3. *If* $k \in \{4, 6, 8, 10, 14\}$, *then* M_k *is one-dimensional, generated by* E_k.
4. $S_k = 0$ *if* $k < 12$ *or* $k = 14$. *Letting* $\Delta := (E_4^3 - E_6^2)/1728 = q - 24q^2 + 252q^3 + \ldots$ *be the* **modular discriminant** *(the name comes from the discriminant of the elliptic curve* $\mathbb{C}/(\mathbb{Z}\tau + \mathbb{Z})$*), we have* $S_{12} = \langle\Delta\rangle$. *For* $k > 14$, $S_k = \Delta M_{k-12}$.

(5) $M_k = S_k \oplus \mathbb{C}E_k$ *for* $k > 2$. *That is, the space of modular forms splits up as cusp forms and Eisenstein series. This will be a powerful fact that's true in great generality. We will see that this is actually an* **orthogonal** *decomposition with respect to a special inner product.*
(6) *The* **algebra** *of modular forms* $\mathcal{M} := \cup_{k\in\mathbb{Z}} M_k$ *is* $\mathbb{C}[E_4, E_6]$. *That is, any modular form is a polynomial in these two functions.*
(7) *There is a basis of* M_k *consisting of forms with* **rational** *Fourier expansions.*
(8) *The dimensions of* M_k *and* S_k *are given by:*

$$\dim_{\mathbb{C}} M_k = \begin{cases} \left\lfloor \frac{k}{12} \right\rfloor & k \equiv 2 \pmod{12}, \\ \left\lfloor \frac{k}{12} \right\rfloor + 1 & k \not\equiv 2 \pmod{12}, \end{cases}$$

$$\dim_{\mathbb{C}} S_k = \dim_{\mathbb{C}} M_k - 1.$$

Proof. We use the valence formula. Since we are looking at holomorphic forms here, all orders of vanishing are non-negative. We look case by case.

Proof of (1): If $f \in M_0$, let c be some value $f(\tau_0)$. Then $f(\tau) - c \in M_0$ has a zero and the valence formula tells us that a bunch of non-negative numbers plus the positive number from the zero at τ_0 add up to zero. This is a contradiction, so in fact $f(\tau) - c \equiv 0$.

Proof of (2): If $k < 0$ or $k = 2$, then we either have a bunch of non-negative numbers adding up to a number that's negative, or we have a sum of integers and positive integers divided by $2, 3$ adding up to $1/6$. This is impossible, as the smallest non-zero sum we could get is $1/3$.

Proof of (3): Let's split into sub-cases.

Case i). $k = 4$ Then the weighted sums of orders of vanishing is $1/3$. The only possibility is that $\nu_\omega(f) = 1$ and there are no other zeros.

Case ii). $k = 6$ Here the weighted sum is $\frac{1}{2}$, so we have $\nu_i(f) = 1$ and there are no other zeros.

Case iii). $k = 8$ In this case, we have $k/12 = 2/3$, so we must have $\nu_\omega(f) = 2$ and there are no other zeros.

Case iv). $k = 10$ Now $k/12 = 5/6$ so there is a simple zero at i and ω and no other zeros.

Case v). $k = 14$ Here we must have $\nu_\omega(f) = 2$, $\nu_i(f) = 1$, and no other zeros.

In any of these cases, if $f_1, f_2 \in M_k$, then they have precisely the same set of zeros. Thus, f_1/f_2 is a holomorphic modular form of weight 0, and by the above constant. Thus, they live in a one-dimensional space, spanned by any non-zero element such as E_k.

Proof of (4): The condition that f is a cusp form means that $\nu_\infty(f) \geq 1$. This can't happen for $k < 12$. If $k = 12$, then there is only one zero allowed, and so f has to have a zero at ∞ and no other zeros. As in the last case, this determines it up to multiples once you find a single example. Δ is such an example as the constant terms of any power of

E_k is 1 so the constant term of $E_4^3 - E_6^2$ is 0. Finally, if $k > 14$ and $f \in S_k$, then we just saw that Δ has no zeros in $\mathbb{H}$, so $f/\Delta \in M_{k-12}$. We'll see a more illuminating proof of the fact that Δ has no zeros on the upper half-plane later.

Proof of (5): Since $E_k = 1 + O(q)$, given any modular form $f \in M_k$, there is a linear combination of f and E_k which kills the constant term of f. This is a cusp form.

Proof of (6): This follows by induction on k. For the one-dimensional spaces $M_4, M_6, M_8, M_{10}, M_{14}$, we can take as generators $E_4, E_6, E_4^2, E4E_6, E_4^2E_6$. Let $k = 12$ or $k > 14$. Then there is a choice of i, j such that $4i + 6j = k$, which gives $E_4^iE_6^j \in M_k$. Since $E_4^iE_6^j = 1 + O(q)$, given any $f \in M_k$, there is a linear combination of f and $E_4^iE_6^j$ which kills the constant term, giving a cusp form. But we saw that cusp forms of these weights are Δ times a form of smaller weight. Since $\Delta \in \mathbb{C}[E_4, E_6]$, by the induction hypothesis we're done.

Proof of (7): The basis is $\{E_4^iE_6^j \mid i, j \geq 0,\ 4i + 6j = k\}$. This is proven by a very similar induction.

Proof of (8): This follows from the proof of the last fact, or another induction. One can prove completely combinatorially the key related fact:

$$\#\left\{i, j \geq 0 \mid 4i + 6j = k\right\} = \begin{cases} \left\lfloor \frac{k}{12} \right\rfloor & k \equiv 2 \pmod{12}, \\ \left\lfloor \frac{k}{12} \right\rfloor + 1 & k \not\equiv 2 \pmod{12}. \end{cases}$$

□

In particular, we obtain a new proof of the convolution sum identities that came from the special differential equation of $\wp(z)$. The recursion we gave for Eisenstein series there will still be useful, but any particular case is a finite check.

Corollary 1.1. For all $n \geq 1$, we have $\sigma_7(n) = \sigma_3(n) + 120 \sum_{0<k<n} \sigma_3(k)\sigma_3(n-k)$.

Proof. Since M_8 is one-dimensional, and $E_4^2, E_8 \in M_8$, we have that $E_4^2 = E_8$ for some constant c. But both have constant term 1, so $c = 1$ and $E_4^2 = E_8$. Now compare Fourier expansions. □

Exercise 1. *Let* $\dim_{\mathbb{C}}(M_k) =: d$. *Show that there is a basis (its called the* **Victor Miller basis***) of* M_k *of elements* $f_0, \ldots, f_{d-1} \in M_k \cap \mathbb{Z}[[q]]$ *(with integer Fourier coefficients) such that for all* $0 \leq i, j \leq d-1$, *the* i*-th coefficient of* f_j, *denoted* $[q^i]f^j$, *is*

$$[q^i]f^j = \delta_{ij},$$

the Kronecker delta function. That is, there is a basis of forms with integer coefficients such that the table of coefficients up to q^{d-1} *is a diagonal of* 1*'s. (Hint: Use row reduction.)*

Exercise 2. *Deduce that two modular forms are the same if and only their Fourier coefficients up to* q^{d-1} *are: a* **finite check***.*