Computing and Verifying Maass Forms

David Lowry-Duda

March 2022 at BYU

ICERM
Brown University

Acknowledgements

This is a project I've begun since joining the Simons Collaboration on Arithmetic Geometry, Number Theory, and Computation. I've collected a large amount of data associated to Maass forms, but there remains a lot to compute and a lot to prove.

In this talk, I'll touch on work done with several collaborators. In particular, I've been working with Andrew Booker (Bristol), Andrei Seymour-Howell (Bristol), and Drew Sutherland (MIT) on computational aspects, and Min Lee, Jonathan Bober, Andrei Seymour-Howell, and Andrew Booker (all at Bristol) with theoretical aspects.

I should also note that I've had the benefit of several helpful conversations with David Farmer (AIM), Sally Koutsoliotas (Bucknell), Stefan Lemurell (Chalmers), Fredrik Strömberg (Nottingham), and the rest of the Simons Collaboration.

Maass forms

Maass forms

Heuristically computing a Maass form

Rigor

Motivation: why study Maass forms?

In his 1966 paper *Can one hear the shape of a drum?*, Mark Kac considered whether the variety of tones and frequencies that can be produced by a drum uniquely identifies the shape of that drum. These tones and frequencies correspond to eigenvectors and eigenvalues of the Laplacian (Helmholtz) equation on drumhead-space.

Maass forms are solutions to the Laplacian differential equation on modular surfaces, as fundamental to modular forms as sound waves are to music.

In practice, Maass forms extend the classical theory of Dirichlet series with Euler products and the theory of classical holomorphic modular forms. The spectral theoretic decomposition into Maass forms led to the discovery of Selberg's trace formula, which connects the spectrum to the underlying geometry.

Personally, I frequently use spectral theory and poor understanding of Maass forms is the most common major obstruction I face.

For this talk, a Maass form will be a *weight* 0 *Maass cuspform* on a congruence subgroup of $\mathrm{SL}(2, \mathbb{Z})$. Specifically, let $\Gamma < \mathrm{SL}(2, \mathbb{Z})$ be a congruence subgroup. The modular surface $X = \Gamma \backslash \mathcal{H}$ is a finite non-compact surface. The Laplacian Δ on this surface is $\Delta = -y^2(\partial^2/\partial x^2 + \partial^2/\partial y^2)$.

We call a function $f : \mathcal{H} \longrightarrow \mathbb{C}$ a Maass cuspform if

1. f is real analytic, $f \in C^\infty(\mathcal{H})$,
2. f is an eigenfunction of the Laplacian, $\Delta f = \lambda f$,
3. f is automorphic, $f(\gamma z) = f(z)$ for all $\gamma \in \Gamma$,
4. f is square integrable, $f \in L^2(X)$, and
5. f vanishes at all the cusps of X.

Selberg famously conjectured that (for congruence subgroups Γ) the eigenvalue λ is either 0 or $\lambda \geq \frac{1}{4}$. An eigenvalue $\lambda \in (0, \frac{1}{4})$ would be called *exceptional*, though we've never seen one.

This *Selberg eigenvalue conjecture* (SEC) is analogous to the Ramanujan–Petersson Conjecture (RPC). We describe this now.

Given a classical (weight k Hecke) holomorphic cusp form

$$g(z) = \sum_{n \geq 1} a(n) n^{\frac{k-1}{2}} e^{2\pi i n z},$$

one can associate an L-function

$$L(s, g) = \sum_{n \geq 1} \frac{a(n)}{n^s} = \prod_p L_p(s),$$

where $L_p(s)$ is (generically) of the form

$$L_p(s) = (1 - \beta_{p,1} p^{-s})^{-1} (1 - \beta_{p,2} p^{-s})^{-1}.$$

The RPC asserts that $|\beta_{p,j}| = 1$, or equivalently that $\log_p |\beta_{p,j}| = 0$.

For holomorphic cusp forms, the RPC is known and follows from Deligne's celebrated proof [Del71].

To each Maass form, there is also an associated L-function. In its completed form, the L-function associated to a Maass form f has the shape

$$\Lambda(s, f) = L_\infty(s) \prod_p L_p(s),$$

where (for generic p)

$$\begin{aligned} L_p(s) &= (1 - \alpha_{p,1} p^{-s})^{-1} (1 - \alpha_{p,2} p^{-s})^{-1} \\ L_\infty(s) &= \Gamma_{\mathbb{R}}(s - \mu_{\infty,1}) \Gamma_{\mathbb{R}}(s - \mu_{\infty,2}). \end{aligned}$$

Here, $L_\infty(s)$ is the "factor at ∞" and consists of a pair of gamma functions $\Gamma_{\mathbb{R}}(s) := \pi^{-s/2}\Gamma(s/2)$.

The parameters $\mu_{\infty,j}$ are closely related to the eigenvalues, and SEC states that $\operatorname{Re} \mu_{\infty,j} = 0$ while RPC states that $\log_p |\alpha_{p,j}| = 0$.

The best progress towards these conjectures for Maass forms are due to Kim and Sarnak, who showed that $|\operatorname{Re} \alpha_{\infty,j}|$ and $\big| \log_p |\alpha_{p,j}| \big|$ are bounded above by $\frac{7}{64}$ [KS03].

Finally, each function $g \in L^2(\Gamma\backslash\mathcal{H})$ has a spectral expansion of the shape

$$\begin{aligned} g(z) = &\sum_{f \text{ Maass cuspform}} \langle g, f\rangle f(z) \\ &+ \sum_{\text{Eisenstein}} \int \langle g, E(\cdot, u)\rangle E(z, u) du \\ &+ (\text{a constant}). \end{aligned}$$

A common *hammer* in my *tool belt* is to average over a family of modular forms, represent everything in terms of the spectral decomposition, and roll up my sleeves and do complex analysis on what remains. The Maass forms that appear in these expansions are typically the barrier to better results.

Maass forms in the LMFDB

The L-function and modular form database (`https://LMFDB.org`) is an online database of L-functions, modular forms, abelian varieties, and their relationships.

There is currently heuristic data for nearly 15000 Maass forms in the LMFDB, available through the portal `https://www.lmfdb.org/ModularForm/GL2/Q/Maass/`.

I'm working on computing more data and making these computations rigorous.

Heuristically computing a Maass form

Maass forms

Heuristically computing a Maass form

Rigor

Methods for the computation of Maass forms have been considered and developed by several authors since the 1970s. Today, I'll describe my preferred method (for GL(2) type Maass forms): Hejhal's algorithm.

In my experience, Hejhal's algorithm is faster and more versatile compared to earlier methods. On the other hand, Hejhal's algorithm is *not rigorous* (although in practice it always produces reliable results). We'll return to the topic of rigorous evaluation later.

The algorithm that Hejhal described apply for the computation of Maass forms for cofinite Fuchsian groups Γ such that $\Gamma\backslash\mathcal{H}$ has exactly one cusp, but I'll also describe the necessary adjustments for when $\Gamma\backslash\mathcal{H}$ has multiple cusps, as is the case for general congruence subgroups Γ.

Maass form Fourier expansion

It is easiest to first describe using Hejhal's algorithm to compute a known Maass newform. Let us fix a Maass form f with eigenvalue $\lambda = \frac{1}{4} + R^2$. Then f has a Fourier expansion

$$f(z) = \sum_{n \neq 0} c(n) \sqrt{y} \frac{W_{iR}(2\pi |n| y)}{\sqrt{n}} e(nx). \tag{1}$$

Here and later, we use the notation $e(nx) = e^{2\pi i n x}$ and $W_{iR}(u) = e^{\pi R/2} \sqrt{u} K_{iR}(u)$, where $K_\alpha(u)$ is the modified K-Bessel function of the second kind.

In this normalization, $W_{iR}(u)$ is an oscillating function of u for $0 < u \lesssim R$ with amplitude roughly of size 1, and then it decays exponentially for $u \gtrsim R$.

In terms of (1), we interpret our goal of *computing a Maass form* to mean to *find the eigenvalue parameter R and the coefficients $c(n)$.*

The coefficients $c(n)$ satisfy the trivial Hecke bound $c(n) = O(\sqrt{n})$ (in fact, much better bounds are known). We can further assume that $c(1) = 1$. Let us fix a desired error bound 10^{-D}. Then there is a decreasing function $M(y) = M(y, R)$ such that

$$f(x + iy) = \sum_{|n| \leq M(y)} c(n)\sqrt{y}\frac{W_{iR}(2\pi|n|y)}{\sqrt{n}}e(nx) + [[10^{-D}]],$$

(where we use $[[10^{-D}]]$ to mean a quantity of absolute value strictly less than 10^{-D}).

Thus we can view $f(x + iy)$ as a finite Fourier series in x up to a small, controlled error.

$$f(x+iy) = \sum_{|n|\leq M(y)} c(n)\sqrt{y}\frac{W_{iR}(2\pi|n|y)}{\sqrt{n}}e(nx) + [[10^{-D}]].$$

Fix a set of equally spaced points along a horocycle

$$\{z_m = x_m + iY : x_m = \frac{1}{2Q}(m - \tfrac{1}{2}), 1 - Q \leq m \leq Q\}$$

(with $Q > M(Y)$). If we think of evaluating f at these points, we are *almost* performing a discrete Fourier transform. Inverting this transform, we see that

$$c(n)\sqrt{Y}\frac{W_{iR}(2\pi|n|Y)}{\sqrt{n}} = \frac{1}{2Q}\sum_{1-Q=m}^{Q} f(z_m)e(-nx_m) + [[10^{-D}]].$$

For fixed R and Y, we can vary n to get essentially a linear system in the coefficients $c(n)$ — but this system is currently a tautology.

We make this system non-tautological by using the automorphy of f, that $f(\gamma z) = z$ for all $\gamma \in \Gamma$. To accomplish this, for the points $z_m = x_m + iY$ in our horocycle, we choose Y small enough so that part of the horocycle will be outside a fixed fundamental domain for $\Gamma\backslash\mathcal{H}$.

Then we pullback each z_m to a point z_m^* in the fundamental domain. The result is that

$$c(n)\sqrt{Y}\frac{W_{iR}(2\pi|n|Y)}{\sqrt{n}} = \frac{1}{2Q}\sum_{1-Q=m}^{Q} f(z_m)e(-nx_m) + [[10^{-D}]].$$

becomes

$$c(n)\sqrt{Y}\frac{W_{iR}(2\pi|n|Y)}{\sqrt{n}} = \frac{1}{2Q}\sum_{1-Q=m}^{Q} f(z_m^*)e(-nx_m) + [[10^{-D}]].$$

If instead of a congruence subgroup, we were considering $\mathrm{SL}(2,\mathbb{Z})\backslash\mathcal{H}$, we would be done. We could expand each $f(z_m^*)$ in its own (essentially finite) Fourier series, repeat for several n, and get a linear system with unknowns $c(n)$. This is the classical algorithm of Hejhal.

Expansions at all the cusps

But when $\Gamma\backslash\mathcal{H}$ has multiple cusps, the resulting linear system is typically very poorly-conditioned. Heuristically this is because several points $z_m = x_m + iY$ might still be in the fundamental domain, and thus $f(z_m) = f(z_m^*)$ for these points — the system is insufficiently mixed by the modularity.

To resolve this, we work not just with the Fourier expansion of f at ∞. We instead work simultaneously with the Fourier expansions f_ℓ at each cusp ℓ. That is, in terms of the Fourier expansions $f_\ell(z) = f(\sigma_\ell z)$, where $\sigma_\ell \infty = \ell$ is a cusp normalization map.

For each point z^* in the fundamental domain, we identify the nearest cusp $\ell = \ell(z^*)$. (By nearest, we mean the cusp with respect to which z^* has the greatest height). Then we represent the value $f(z^*)$ in terms of the Fourier expansion f_ℓ.

(This is the lots-of-bookkeeping aspect of the approach). In order to set up the extended system, we must enlarge our linear system to include horocycles associated to the expansion at each cusp and solve for all expansions simultaneously. For each cusp j, we have an expansion

$$f_j(z) = \sum_{n \neq 0} c_j(n) \sqrt{y} \frac{W_{iR}(2\pi |n| y)}{\sqrt{n}} e(nx)$$

and we can set up the system

$$c_j(n) \sqrt{Y} \frac{W_{iR}(2\pi |n| Y)}{\sqrt{n}} = \frac{1}{2Q} \sum_{1-Q=m}^{Q} f_j(z_m) e(-nx_m) + [[10^{-D}]]$$

as before.

We now have the system

$$c_j(n)\sqrt{Y}\frac{W_{iR}(2\pi|n|Y)}{\sqrt{n}} = \frac{1}{2Q}\sum_{1-Q=m}^{Q} f_j(z_m)e(-nx_m) + [[10^{-D}]].$$

Let $z_{mj} = \sigma_j z_m$, so that $f_j(z_m) = f(z_{mj})$, and let z^*_{mj} be the pullback of z_{mj} to the fundamental domain, expressed in coordinates of the nearest cusp ℓ. Automorphy implies that $f(z_{mj}) = f_\ell(z^*_{mj})$, and in total

$$c_j(n)\sqrt{Y}\frac{W_{iR}(2\pi|n|Y)}{\sqrt{n}} = \frac{1}{2Q}\sum_{1-Q=m}^{Q} f_\ell(z^*_{mj})e(-nx_m) + [[10^{-D}]].$$

Lemma
It is possible to choose Y small enough such that $z^*_{mj} \neq z_{mj}$ for all j and m. Further, the imaginary parts of each resulting z^*_{mj} are bounded below by a computable constant Y_0 (which depends on the level of the congruence subgroup).

It is the nontrivial mixing coming from $f_j(z_m)$ and $f_\ell(z^*_{mj})$ that gives a non-tautological system, allowing us to solve for the Fourier coefficients in the linear system.

Solving for the coefficients

Summarizing so far: given an input eigenvalue $\lambda = \frac{1}{4} + R^2$, we can set up the system

$$c_j(n)\sqrt{Y}\frac{W_{iR}(2\pi|n|Y)}{\sqrt{n}} = \frac{1}{2Q}\sum_{1-Q=m}^{Q} f_\ell(z^*_{mj})e(-nx_m) + [[10^{-D}]].$$

If we choose the Y in the horocycles as in the Lemma, then $\mathrm{Im}(z^*_{mj}) > Y_0$ for all m and j, so we can truncate each Fourier series f_ℓ on the right at the same point $M_0 = M(Y_0)$ while guaranteeing a uniform error bound. Expanding each finite Fourier series and collecting coefficients, we get that

$$c_j(n)\sqrt{Y}\frac{W_{iR}(2\pi|n|Y)}{\sqrt{n}} = \sum_{\text{cusps } \ell}\sum_{1\le|k|\le M_0} c_j(k)V_{nkj\ell} + 2[[10^{-D}]]$$

for complicated-but-computable coefficients $V_{nkj\ell}$ (that are just complicated combinations of K-Bessel functions and exponentials). Considering this for each $|n| \le M_0$ gives a linear system that can be solved.

Structurally, we have constructed a homogeneous linear system $V\vec{c} = 0$ for a computable matrix $V = V(R, Y)$ consisting mostly of linear combinations of Bessel functions and an unknown vector of coefficients $\vec{c}$.

We can use the assumption $c(1) = 1$ to de-homogenize the linear system and to facilitate solving for the coefficients.

It should be noted that a priori, it is not obvious that the resulting linear system will be well-conditioned. This would be a necessary ingredient to conclude that this algorithm would always succeed, but this is unknown. However, in my experiments it seems that whenever we choose Y small enough so that $z_{mj} \neq z_{mj}^*$ for all m and j, the resulting system is solvable and gives approximately D correct digits of accuracy for the coefficients.

Dimension reduction

There are frequently relations between the cusps that allow one to reduce the dimension of the linear system. In particular, there are Hecke-operator type symmetries (Fricke involutions) that connect Fourier expansions at cusps.

I'll also remark that all the work here carries through even when there is a nontrivial nebentypus, except that one must track the character and how it carries through the cusp-normalizing maps σ_ℓ. (This is simply additional bookwork).

Rigor

Maass forms

Heuristically computing a Maass form

Rigor

We have demonstrated that we can heuristically determine a Maass form with a known eigenvalue by constructing a homogeneous linear system $V\vec{c} = 0$.

To make this rigorous, we need some way to determine the eigenvalue R, and we need some way to verify that our claimed Maass form is *close* to a true Maass form.

Isolating eigenvalues

The Selberg Trace formula relates eigenvalues of Maass forms to the geometry of the group, giving a relation (loosely) of the form

$$\sum h(r_n) = \int_{-\infty}^{\infty} rh(r) \tanh(\pi r) dr + \sum_{\text{conj } T \in \Gamma} (*) \widetilde{h}(T).$$

In forthcoming work, Andrei Seymour-Howell describes how to explicitly implement the Selberg Trace formula to get low quality estimates for eigenvalues.

In practice, we'll start with these *approximations* to eigenvalues and use Hejhal's algorithm to improve the estimates.

Verification

To understand how to perform rigorous verification, we need to examine the sources of error in the linear system.

Recall that we have a homogenized linear system $V(r)\vec{c} \approx 0$, which is a $(\kappa M) \times (\kappa M)$ dimensional system, where κ is the number of distinct cusps and M is the number of Fourier coefficients we use for truncated Fourier series.

Using $c(1) = 1$, we obsolete the first column of the matrix and remove the first row to think of as an auxiliary equation. We then have a $(\kappa M - 1) \times (\kappa M - 1)$ dimensional system $V(r)\vec{c} \approx b(r)$, and an auxiliary equation of the form $\vec{c} \cdot v(r) + w(r) \approx 0$.

(One form of Hejhal's algorithm is to iteratively solve $V(r)\vec{c} = b(r)$ while minimizing the error from the auxiliary equation $A(r) := \vec{c} \cdot v(r) + w(r)$).

Idea of verification

Suppose that the interval $[r - \epsilon, r + \epsilon]$ is known to contain a unique eigenvalue R. Write $R = r + \delta$.

We can bound the error δ numerically using the auxiliary equation $A(r) = \vec{c} \cdot v(r) + w(r)$.

From the Taylor expansion

$$A(R) = A(r) + A'(r)\delta + A''(\widetilde{r})\delta^2/2$$

(which holds for some $\widetilde{r}$ between r and R), we find that

$$|\delta| = \frac{|A(r) - A(R)|}{|A'(r) + A''(\widetilde{r})\delta/2|}.$$

When the initial guess r is good enough, the numerator should be very small and the denominator is approximately $A'(r)$. Thus we approximate $A'(r)$ very carefully and bound everything else.

$$|\delta| = \frac{|A(r) - A(R)|}{|A'(r) + A''(\widetilde{r})\delta/2|}.$$

It's concerning that R (which we don't know) appears in this equation. But $A(R)$ is the auxiliary equation coming from the first row of the linear system, and should be very *very* small. The only reason why it's not zero is due to error from truncating Fourier expansions. Thus even though it's not known, it can be bounded.

Other Complications

1. The eigenvalue candidate r is only *near* a true eigenvalue R, and we don't know R.
2. We also don't know the precise value of $\widetilde{r}$ in the Taylor expansion above.
3. Truncating the Fourier expansions of $f(z)$ gives many small errors.
4. In all computations, we work with Bessel functions and their derivatives, which accrue computational errors.

And in the end, it's possible that this verification algorithm might fail, indicating that our initial estimate wasn't good enough.

But when it does work, it works very well. Iterating produces even better estimates.

Thank you very much.

Please note that these slides (and references for the cited works) are (or will soon be) available on my website (davidlowryduda.com).

References i

P. Deligne.
Formes modulaires et représentations l-adiques.
Séminaire N. Bourbaki, **355**:139–172, 1971.

H. Kim and P. Sarnak.
Refined estimates towards the Ramanujan and Selberg conjectures.
J. Amer. Math. Soc., **16**:175–181, 2003.