Rigorously computing Maass forms

David Lowry-Duda

February 2023

Concordia University
QVNTS

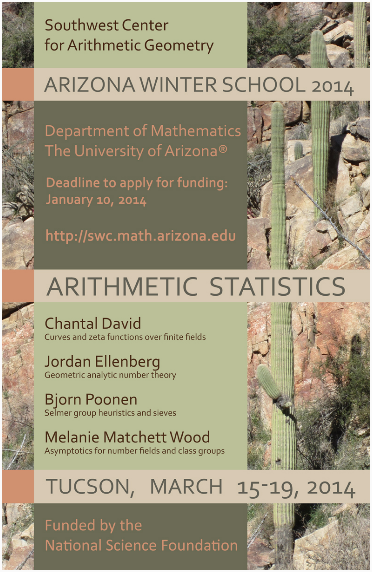
This is a project I've begun since joining the Simons Collaboration on Arithmetic Geometry, Number Theory, and Computation. I've collected a large amount of data associated to Maass forms, but there remains a lot to compute and a lot to prove.

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 with Min Lee (Bristol) on 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.

I'd like to thank Chantal
and the organizers for inviting me.

One of my first
papers was with Chantal, coming
out of her project group at
the 2014 Arizona Winter School!



Motivation and Context

The geometry of a space strongly influences the functions that live on that space. For a familiar example, consider drums.

The shape of a drumhead affects the sounds that a can make. The frequencies at which a drumhead can vibrate are determined by the Helmholtz equation,

$$\begin{cases} \Delta u + \lambda u = 0, \\ u_{\partial D} = 0. \end{cases}$$

Different drums admit different solutions, having different tones.

Aside: it is also interesting to ask,

Can one can hear the shape of a drum?

Answer: no!

Maass forms are solutions to a Laplacian differential equation with a certain boundary (cusp) condition on modular surfaces, and are as fundamental to modular forms as sound waves are to music.

But unlike sound waves or drum tones, Maass forms are not simple. They're extremely mysterious and enigmatic.

We will see that Maass forms extend the classical theory of Dirichlet series with Euler products and the theory of classical holomorphic modular forms.

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.

History

We call these Maass forms because Maass was the first to look for them. Maass was trying to extend Siegel's work on theta functions and found some evidence that an unknown family of functions might exist. Rephrasing to look more like what Maass wrote (and more like the Helmholtz equations for drums), Maass looked for solutions to

$$\begin{cases} \Delta f - \lambda f = 0, \\ f(\gamma z) = f(z), \\ f \in L^2(\Gamma \backslash \mathcal{H}). \end{cases}$$

Maass showed that for any fixed λ, there are at most finitely many solutions, but he was unsuccessful in finding any outside of very particular constructions (essentially coming from Hecke characters).

History

In 1956, Selberg developed what we now call the *Selberg trace formula* and the theory of real analytic Eisenstein series in order to prove the general existence of Maass forms on $\mathrm{SL}(2, \mathbb{Z})$.

This has had many far reaching implications!

The Selberg trace formula was generalized to the Arthur-Selberg trace formula, used by Jacquet, Langlands, and many others to prove special cases of Langlands functoriality.

But so far, no one has ever exactly computed a Maass form on $\mathrm{SL}(2, \mathbb{Z})$. It could be that they're uncomputable!

Before describing how we're going to try to compute Maass forms, I'd like to give a bit more context.

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 modular 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)

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

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. The 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].

I first began to investigate Maass forms because they kept on appearing in my number theoretic results. Continuing to use a weight k holomorphic cuspform

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

we can look at the convolution L-function

$$L(s, g \otimes g) = \zeta(2s) \sum_{n \geq 1} \frac{a(n)^2}{n^s}.$$

This has a functional equation, Euler product, meromorphic continuation — it's a very natural L-function.

The convolution series

$$D_h(s) = \sum_{n \geq 1} \frac{a(n)a(n+h)}{n^s}$$

doesn't have a functional equation or Euler product, but it does have a meromorphic continuation to $\mathbb{C}$ with poles at $\frac{1}{2} \pm it_j$ for each eigenvalue $\lambda_j = \frac{1}{4} + t_j^2$ of a Maass form on $\Gamma(N)$.

Similar results hold even for objects that *seem* of independent arithmetic interest. Let $d(n) = \sum_{\ell | n} 1$ denote the standard divisor function. Then one can also show that

$$\sum_{n \geq 1} \frac{d(n)d(n+h)}{n^s}$$

has meromorphic continuation to $\mathbb{C}$ with poles coming from each eigenvalue of Maass forms on $\mathrm{SL}(2, \mathbb{Z})$.

I'm interested in the Gauss circle problem and its generalizations. This concerns understanding the discrepancy between the number of lattice points $S_2(R)$ in a circle of radius $\sqrt{R}$ and the "obvious" main term πR,

$$P_2(R) = S_2(R) - \pi R.$$

In [HKLDW21], Hulse, Kuan, Walker, and I looked at the Laplace transform of $P_2(R)^2$ and found that

$$\begin{aligned}\int_0^\infty P_2(t)^2 e^{-t/X} dt =& cX^{\frac{3}{2}} + c_1 X + \sum c_{t_j} X^{\frac{1}{2}+it_j} \\ &+ \sum c_\rho X^\rho + O(X^{\frac{1}{4}+\epsilon}).\end{aligned}$$

Here, $\frac{1}{2} + it_j$ again comes from Maass forms, and ρ comes from zeros of $\zeta(2s)$. Understanding $P_2(R)$ is hard because it is coupled to the distribution of Maass eigenvalues.

A huge number of Diophantine problems and problems in arithmetic are controlled by the distribution of Maass forms.

Spectral expansion

One reason why is that each function $g \in L^2(\Gamma \backslash \mathcal{H})$ has a spectral expansion of the shape

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

Automorphic forms appear all over, and their analytic behavior can be understood in terms of their spectra.

Computing Maass forms

Our goal is to rigorously compute Maass forms. The problem is that everything associated to a generic Maass form is transcendental and one will never *exactly* compute a Maass form.

Each Maass form discussed today has an expansion

$$f(z) = \sum_{n \geq 1} \frac{a(m)}{\sqrt{m}} W_{ir}(2\pi m y) e(2\pi m x),$$

This is a (real) analytic function on $\Gamma_0(N)\backslash\mathcal{H}$ for some squarefree N, and it is an eigenfunction of a Laplacian with eigenvalue $\lambda = \frac{1}{4} + r^2$.

By "compute a Maass form", we mean to rigorously estimate the eigenvalue λ (or equivalently the spectral parameter r) and to rigorously estimate the coefficients $a(m)$.

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 thousands of Maass forms in the LMFDB, available through the portal `https://www.lmfdb.org/ModularForm/GL2/Q/Maass/`.

In the next couple of months, I will upload more data, with rigorous results.

Heuristically computing Maass forms

We base our method for computation on an algorithm Hejhal developed in the 1970s to find a Maass form.

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

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

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 $u \gtrsim R$.

Thus we want to understand R and the coefficients $c(n)$.

The coefficients $c(n)$ satisfy the trivial Hecke bound $c(n) = O(\sqrt{n})$ (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}).

We can think of $f(x + iy)$ 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.

In order to set up the extended system, we 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

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

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

Structurally, we have constructed a homogeneous linear system $T\vec{c} = 0$ for a computable matrix $T = T(R, Y)$, 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 practice 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.

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

But in practice, we don't know the eigenvalue R. Hejhal also gave a method to try to find R.

It is an interesting fact that we don't know a good way to strongly approximate R without simultaneously finding strong approximations for the coefficients.

Abstractly, we can think of our linear system as being of the form

$$T(r)\vec{a} \approx 0,$$

where $\vec{a} = \big(a(k)\big)_{1 \le k \le L}^{T}$ and we feed in r as an input.

When we use $a(1) = 1$ to dehomogenize the system, we remove first column of the system of equations (corresponding to $a(1)$) and separate the first row as an auxiliary equation. Explicitly (and abusing notation), we have

$$T(r)\vec{a} \approx b(r),$$

where now $\vec{a} = \big(a(k)\big)_{2 \le k \le L}^{T}$ and $b(r)$ are explicit in terms of the coefficients of $a(1) = 1$.

The auxiliary equation from the first row can be written

$$c(r) := \vec{a} \cdot v(r) + w(r) \approx 0.$$

Iteration

$$T(r)\vec{a} \approx b(r), \tag{1}$$

$$c(r) := \vec{a} \cdot v(r) + w(r) \approx 0. \tag{2}$$

Note that the matrix $T(r)$ and components $b(r), v(r)$, and $w(r)$ depend on the (a priori unknown) parameter r. One form of Hejhal's algorithm is to guess an initial r, solve (1) to get approximations to the coefficients $a(k)$, and then iterate while trying to minimize the error term in the auxiliary equation (2).

One part of making this rigorous is to find precise error bounds in the linear system (1) and the first coefficient auxiliary equation (2), including the error coming from truncation.

Rigorously computing Maass forms

The first (and only) rigorously computed Maass forms were done by Booker, Strömbergsson, and Venkatesh in 2006 [BSV06]. There, they compute the first dozen or so Maass forms on $\mathrm{SL}(2,\mathbb{Z})$ to over a thousand digits of precision.

They probed the algebraic and transcendental properties of the coefficients and the eigenvalue — they seem transcendental.

But their method is specialized to $\mathrm{SL}(2,\mathbb{Z})$ and is computationally expensive.

The dream is to have something that is both as rapidly computable as Hejhal's algorithm *and* rigorous. For this, we need an additional ingredient.

The key difference from my original efforts to find Maass forms has been the unexpectedly strong success of a method to find low-precision estimates for the eigenvalues λ using an explicit, computational form of Selberg's trace formula. This is the topic of Andrei Seymour-Howell's PhD thesis.[1]

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

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

Andri takes combinations of specially chosen test functions to find rough approximations for eigenvalues.

[1] Andrei is a student of Andy Booker, and the possibility of this approach was noted in [BSV06].

The broad strategy is now to first compute several intervals $[r - \epsilon, r + \epsilon]$ that are known to contain a unique eigenvalue parameter r^* (with explicit ϵ bound); we then use a rigorous version of Hejhal's algorithm to refine the intervals.

Aside

Andrei's work currently applies only to squarefree level and trivial inner character. In a closely related project, we[2] are working to develop an explicit, computational trace formula for squarefull level and nontrivial character.

[2] also with Bober, Booker, Knightley, and Min Lee

Making Hejhal's algorithm rigorous

Removing $\approx$ signs

We now fix a single interval $[r-\epsilon, r+\epsilon]$ known to contain a unique (but unknown) eigenvalue r^*.

Write $r^* = r + \delta$.

To make things rigorous, we must remove $\approx$ signs from our Hejhal system

$$\begin{aligned} T(r)\vec{a} &\approx b(r), \\ c(r) := \vec{a} \cdot v(r) + w(r) &\approx 0. \end{aligned}$$

To do that, let $b^\natural(r)$ denote the vector we get for r by truncating all Fourier series at L, setting up the system, and ignoring all truncation error terms. Then define

$$e = T(r^*)\vec{a} - b^\natural(r^*), \qquad b(r) := b^\natural(r) + e.$$

Now $b(r)$ is precisely defined (though we are ignorant of its exact value).

Let $\vec{a}(r)$ denote the computed solution for $\vec{a}$ at r in the dehomogenized, now well-defined Hejhal system

$$T(r)\vec{a} = b(r).$$

In practice, we prove bounds using this idealized form but compute in interval arithmetic. Though we don't know $b(r)$ exactly (because we don't know the error e exactly), we can bound the error e. With these definitions, it follows that $\vec{a}(r^*)$ is an *exact* solution for the coefficients of the desired Maass form, and thus the error e comes entirely from truncation error.

Bounding the sizes of the tails of the Fourier expansions for f allows us to compute interval estimates for $b(r)$, and thus interval estimates for $\vec{a}(r)$.

Core Idea

We are now ready to state the core idea of how to refine the error. Recall that the true eigenvalue parameter $r^* \in [r - \epsilon, r + \epsilon]$ and $r^* = r + \delta$. We compute (interval estimates) for $\vec{a}(r)$ and look at the auxiliary system

$$c(r) := \vec{a}(r) \cdot v(r) + w(r).$$

Near r^*, $c(r)$ has the expansion $c(r^*) = c(r) + c'(r)\delta + c''(\widetilde{r})\delta^2/2$ for some $\widetilde{r}$ between r and r^*. Rearranging, we find that

$$|\delta| = \frac{|c(r) - c(r^*)|}{|c'(r) + c''(\widetilde{r})\delta/2|}.$$

The core idea is to find tight (interval) approximations for $c(r)$ and $c'(r)$ and rigorously bound $c(r^*)$ and c''. As $|\delta| \leq \epsilon$, we find for example that if $\epsilon|c''(\widetilde{r})| < |c'(r)|$, then

$$|\delta| \leq 2\left|\frac{c(r) - c(r^*)}{c'(r)}\right| \approx \frac{\text{extremely small}}{c'(r^*)}.$$

If ϵ is small enough in terms of $c'(r^*)$ and $c''(r^*)$, then this strategy succeeds.

But this doesn't always succeed. In the worst-case scenario, we might have $c'(r^*) = 0$. I have never observed this to occur, but if it did then we would have to construct a different auxiliary equation and work from there.

More often, the problem is that our initial ϵ-interval estimate from the trace formula isn't quite strong enough. It is possible to work harder to produce better initial intervals, but there is more work to be done. This is a *later* project.

For the remainder of this talk, I'll comment on estimating and bounding the components of the auxiliary expansion

$$c(r^*) = c(r) + c'(r)\delta + c''(\widetilde{r})\delta^2/2.$$

$c(r)$ and $c(r^*)$

Estimating $c(r)$ is "easy" as long as we can compute $T(r), b^\natural(r), v(r)$, and $w(r)$ very precisely. As previously noted, we can bound the truncation error e arbitrarily well by using enough coefficients, giving an arbitrarily good interval bound for $b(r)$. As long as $T(r)$ is invertible,[3] we compute interval bounds for $\vec{a}(r)$. These directly give interval bounds for $c(r)$.

At the eigenvalue r^*, we should expect $c(r^*) \approx 0$. The only error is also due to truncation error. We can bound this by essentially the same techniques as bounding the truncation error e for $b^\natural(r)$.

[3] If $T(r)$ is not invertible, then Hejhal's algorithm fails and we choose a different horocycle.

$c''(r)$

Bounding $c''(r)$ is the most technically involved aspect of the analysis. A major problem is that naive derivatives here involve derivatives of the Whittaker functions W_{ir}, for which we don't have particularly accurate estimates.[4]

Generally, we can approximate W_{ir} very well, but we can only efficiently compute rigorous bounds for general $\frac{\partial}{\partial r} W_{ir}$ and $\frac{\partial^2}{\partial r^2} W_{ir}$.

We must be careful about which terms we can approximate and which terms we can merely bound.

[4] A surprisingly large amount of my time on Maass forms has been centered on producing various rigorous estimates for Bessel functions.

Taylor-type approximations

For any matrix valued function $M(t)$ that is twice differentiable in a neighborhood of r, we define

$$M_1(r,\delta) := \begin{cases} \frac{M(r+\delta)-M(r)}{\delta} & \delta \neq 0, \\ M'(r) & \delta = 0, \end{cases}$$

$$M_2(r,\delta) := \begin{cases} \frac{M(r+\delta)-M(r)-M'(r)\delta}{\delta^2} & \delta \neq 0, \\ \frac{M''(r)}{2} & \delta = 0. \end{cases}$$

This implies that

$$M(r+\delta) = M(r) + M_1(r,\delta)\delta = M(r) + M'(r)\delta + M_2(r,\delta)\delta^2$$

for sufficiently small δ.

To bound c'', we actually bound $c_2(r,\delta)$.

Second order Hejhal approximation

The Hejhal system $T(t) \cdot \vec{a} = b(t)$ is smooth as W_{it} is smooth, and has first order version

$$(T(r) + T_1(r,\delta)\delta)(\vec{a}(r) + \vec{a}_1(r,\delta)\delta) = b(r) + b_1(r,\delta)\delta,$$

which implies that

$$T(r)\vec{a}_1(r,\delta) = b_1(r,\delta) - T_1(r,\delta)\vec{a}(r+\delta).$$

Fixing again δ so that $r^* = r + \delta$ (and performing similar analysis on the second order version) shows that

$$\begin{aligned}
\vec{a}_1(r,\delta) &= T^{-1}(r)(b_1(r,\delta) - T_1(r,\delta)\vec{a}(r^*)) \\
\vec{a}_2(r,\delta) &= T^{-1}(r)(b_2(r,\delta) - T'(r)\vec{a}_1(r,\delta) - T_2(r,\delta)\vec{a}(r^*)).
\end{aligned}$$

With bounds for the first and partial r-derivatives of $W_{ir}(y)$, we can bound $T'(r)$, $T_1(r,\delta)$, $T_2(r,\delta)$, $b_1(r,\delta)$, and $b_2(r,\delta)$. We bound the exact coefficients $\vec{a}(r^*)$ using the Kim–Sarnak bound. Thus we can bound $\vec{a}_1$ and $\vec{a}_2$.

Similar analysis on the second order version, as well as the first and second order versions of the auxiliary equation, ultimately show that

$$c_2(r,\delta) = w_2(r,\delta) + \vec{a}_2(r,\delta) \cdot v(r) + \vec{a}_1(r,\delta) \cdot v'(r) + \vec{a}(r+\delta) \cdot v_2(r,\delta).$$

Using the bounds for $\vec{a}_1$ and $\vec{a}_2$ from before, we can bound $c_2(r,\delta)$.

$c'(r)$

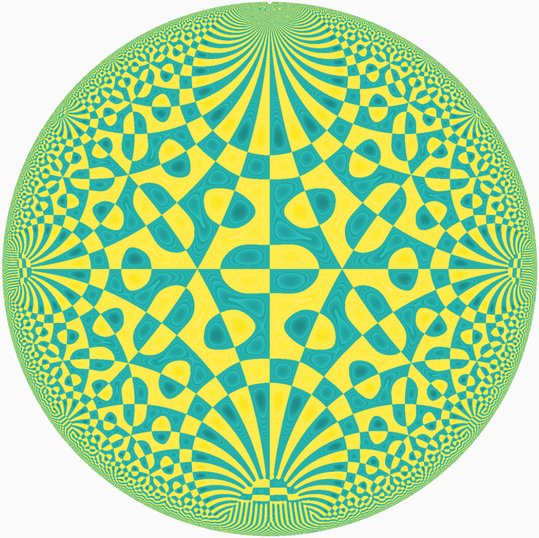
Finally, we approximate $c'(r)$. In principle, we could do this through a first order approximation of the Hejhal system, as we did with c_2 above. But this would require a high accuracy algorithm to compute (not just bound) $\partial_r W_{ir}$.

Instead, we take a first order approximation to $c'(r)$ and apply our bound for c_2, via

$$c'(r) = \frac{c(r+\delta_0) - c(r)}{\delta_0} + c_2(r, \delta_0)\delta_0.$$

We precisely approximated $c(r)$ by solving Hejhal's linear system, and we do this again by solving another linear system for $c(r+\delta_0)$ for a small δ_0.

The final term $c_2(r, \delta_0)$ can be bounded as before, except that instead of using the Kim-Sarnak bound for idealized coefficients $\vec{a}(r^*)$, we use the fact that we now have explicitly computed coefficients $\vec{a}(r+\delta_0)$ and use their actual values.



Thank you very much.

There will be more details in a forthcoming preprint (joint with Andrei Seymour-Howell), and data on an LMFDB near you.

Please note that these slides are (or will soon be) available on my website (davidlowryduda.com).

References i

Andrew R Booker, Andreas Strömbergsson, and Akshay Venkatesh.
Effective computation of maass cusp forms.
International mathematics research notices, 2006.

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

Thomas A. Hulse, Chan Ieong Kuan, David Lowry-Duda, and Alexander Walker.
The Laplace transform of the second moment in the Gauss circle problem.
Algebra Number Theory, 15(1):1–27, 2021.
`https://arxiv.org/abs/1705.04771`.

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