Spectral Theory on Hyperbolic Surfaces

David Borthwick
Emory University
July, 2010

Outline

I. Hyperbolic Geometry

(Escher/Jos Leys)

Möbius transformations

The upper half-plane $\mathbb{H} = \{\operatorname{Im} z > 0\}$ has a large group of conformal automorphisms, consisting of Möbius transformations of the form

$$z \mapsto \frac{az+b}{cz+d},$$

where $a, b, c, d \in \mathbb{R}$ and $ad - bc > 0$.

These symmetries form the group

$$\mathrm{PSL}(2,\mathbb{R}) := \mathrm{SL}(2,\mathbb{R})/\{\pm I\}$$

Hyperbolic metric

Under the $\mathrm{PSL}(2,\mathbb{R})$ action, $\mathbb{H}$ has an invariant metric,

$$ds^2 = \frac{dx^2 + dy^2}{y^2},$$

often called the *Poincaré metric*.

This metric is *hyperbolic*, meaning that the Gauss curvature is -1.

We have a corresponding measure

$$dA(z) = \frac{dx\,dy}{y^2},$$

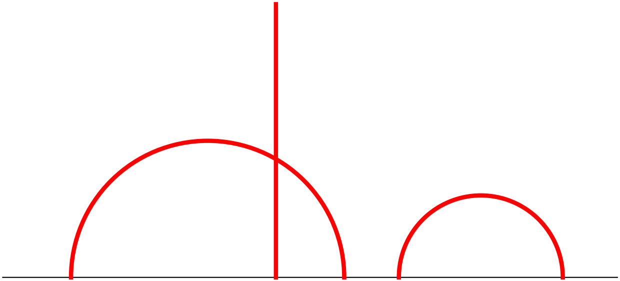
and distance function

$$d(z,w) = \log \frac{|z-\bar{w}| + |z-w|}{|z-\bar{w}| - |z-w|}$$

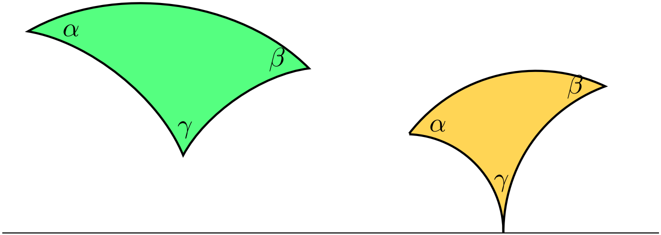
Geometry in $\mathbb{H}$

The hyperbolic metric is conformal to the Euclidean metric, so angles are computed as in Euclidean geometry.

Geodesics are arcs of generalized circles intersecting $\partial\mathbb{H} := \mathbb{R} \cup \{\infty\}$ orthogonally.



The Gauss-Bonnet theorem gives a formula for areas of triangles with geodesic sides.



$$\text{Area}(ABC) = \pi - \alpha - \beta - \gamma.$$

Fixed points

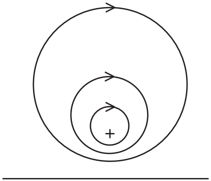
Elements of $\mathrm{PSL}(2,\mathbb{R})$ are classified according to their fixed points.

For $g \in \mathrm{PSL}(2,\mathbb{R})$, the fixed point equation $z = gz$ is quadratic:

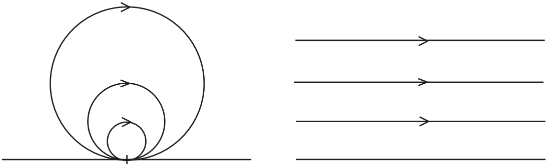
$$cz^2 + (d-a)z - b = 0$$

For each g there are exactly 2 solutions in $\mathbb{C} \cup \infty$.

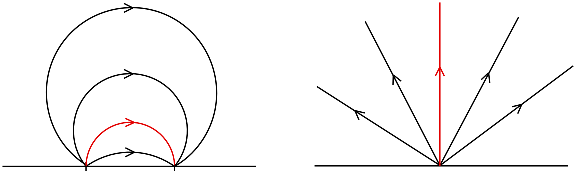
1. *elliptic*: one fixed point in $\mathbb{H}$. (The other must be the complex conjugate.) An elliptic transformation is a rotation centered at the fixed point.



2. *parabolic*: a single degenerate fixed point (must lie on $\partial\mathbb{H}$.) Any parabolic transformation is conjugate to the map $z \mapsto z + 1$.



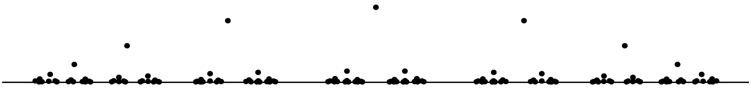
3. *hyperbolic*: two distinct fixed points in $\partial\mathbb{H}$. A hyperbolic transformation is conjugate to the map $z \mapsto e^{\ell} z$ for some $\ell \in \mathbb{R}$.



II. Fuchsian groups

A *Fuchsian group* is a discrete subgroup $\Gamma \subset \mathrm{PSL}(2, \mathbb{R})$ ("discrete" in the matrix topology, equivalent to Euclidean $\mathbb{R}^4$.)

This is equivalent to the condition that Γ acts properly discontinuously on $\mathbb{H}$, meaning that each orbit Γz are locally finite.

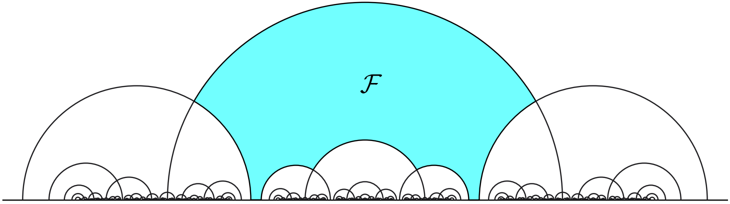


Quotients

The quotient $\Gamma \backslash \mathbb{H}$ is a surface whose points correspond to the orbits of Γ in $\mathbb{H}$. For Fuchsian Γ, the quotient inherits a hyperbolic metric from $\mathbb{H}$.

In general, $\Gamma \backslash \mathbb{H}$ is an *orbifold*. The quotient is a smooth surface iff Γ has no elliptic elements.

A *fundamental domain* for Γ is a closed region $\mathcal{F}$ such that the translates of $\mathcal{F}$ under Γ tesselate $\mathbb{H}$.



Automorphic forms

A function f on $\Gamma\backslash\mathbb{H}$ is equivalent to a function on $\mathbb{H}$ satisfying

$$f(gz) = f(z) \text{ for } g \in \Gamma.$$

The latter is called an *automorphic function* for Γ.

An *automorphic form of weight k* satisfies

$$f(gz) = (cz + d)^k f(z).$$

(These are sections of the k-th power of the canonical line bundle over $\Gamma\backslash\mathbb{H}$.)

Warning: in some contexts automorphic forms are required to be meromorphic or even entire.

Uniformization

[Koebe & Poincaré, 1907]: For any smooth complete Riemannian metric on a surface, there is a conformally related metric of constant curvature. (By scaling, we can restrict this curvature to the values $-1, 0, 1$.)

Up to isometry, the only simply connected possibilities are the sphere S^2, Euclidean $\mathbb{R}^2$, and the hyperbolic plane $\mathbb{H}^2$.

All of the surfaces with Euler characteristic $\chi < 0$ (i.e. most surfaces) are uniformized by hyperbolic quotients $\Gamma\backslash\mathbb{H}$.

Riemann surfaces

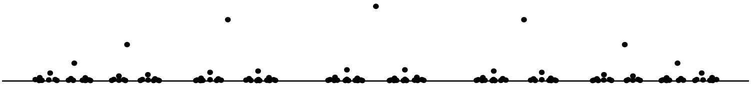
The term *hyperbolic* refers to curvature $= -1$. But because the hyperbolic isometries of $\mathbb{H}$ are the same as the conformal automorphisms, any quotient $\Gamma\backslash\mathbb{H}$ has a natural complex structure.

A *Riemann surface* is a one-dimensional complex manifold.

Uniformization implies that any Riemann surface can be realized as a quotient of the Riemann sphere $\mathbb{C} \cup \{\infty\}$, the complex plane $\mathbb{C}$, and the upper half-plane $\mathbb{H}$.

Limit set

The limit set $\Lambda(\Gamma)$ of $\Gamma \subset \mathrm{PSL}(2, \mathbb{R})$ is the collection of limit points of orbits of Γ, in the Riemann sphere topology.



In fact, it suffices to take the orbit of any single point that's not an elliptic fixed point.

$\Lambda(\Gamma)$ is a closed, Γ-invariant subset of $\partial\mathbb{H} = \mathbb{R} \cup \{\infty\}$.

Classification of Fuchsian groups

[Poincaré, Fricke-Klein]: Fuchsian groups are classified as

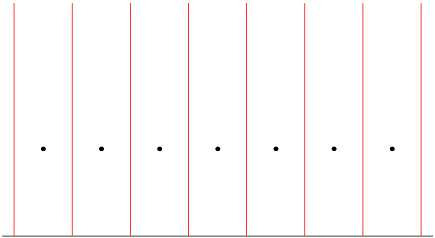
1. *Elementary*: $\Lambda(\Gamma)$ contains 0, 1, or 2 points;
2. *First Kind*: $\Lambda(\Gamma) = \partial\mathbb{H}$;
3. *Second Kind*: $\Lambda(\Gamma)$ is a perfect, nowhere-dense subset of $\partial\mathbb{H}$.

It turns out that

$$\Gamma \text{ is of the first kind} \quad \Longleftrightarrow \quad \mathrm{Area}(\Gamma\backslash\mathbb{H}) < \infty.$$

All arithmetic surfaces are in this category.

The cyclic group $\Gamma_\infty = \langle z \mapsto z + 1 \rangle$ is an example of an elementary group.



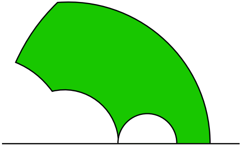
The orbits accumulate at $\Lambda(\Gamma) = \{\infty\}$.

Similarly, the hyperbolic cyclic group $\langle z \mapsto e^\ell z \rangle$ has $\Lambda(\Gamma) = \{0, \infty\}$.

All other elementary groups are finite elliptic groups.

Geometric finiteness

A Fuchsian group is *geometrically finite* if Γ admits a fundamental domain that is a finite-sided convex polygon.



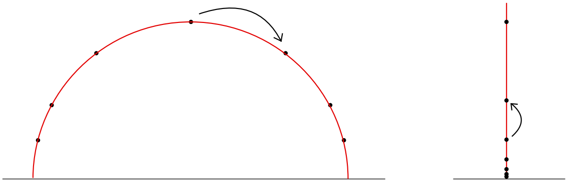
This coincides with two other notions:

1. topological finiteness of $\Gamma\backslash\mathbb{H}$, and
2. Γ is finitely generated.

Spectral theory is only tractable in general for geometrically finite Γ. A theorem of Siegel says that all groups of the first kind are geometrically finite.

Geodesics on $\Gamma\backslash\mathbb{H}$

If $g \in \Gamma$ is hyperbolic, then there is a unique geodesic connecting the two fixed points of g, called the *axis* of g. On its axis, g acts by translation by some fixed length ℓ.

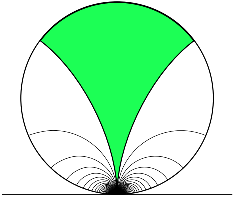


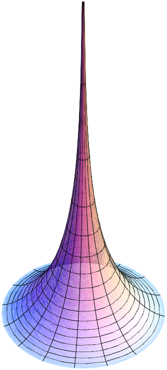
In $\Gamma\backslash\mathbb{H}$, the axis descends to a closed geodesic of length ℓ. There is a 1-1 correspondence:

$$\text{closed geodesics} \longleftrightarrow \text{conjugacy classes of hyperbolic elements.}$$

Cusps

The parabolic elements of Γ create *cusps* in $\Gamma\backslash\mathbb{H}$.



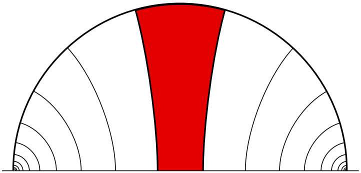


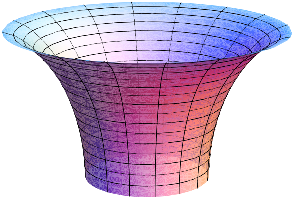
There is a 1-1 correspondence:

$$\text{cusps} \longleftrightarrow \text{orbits of parabolic fixed points.}$$

Funnels

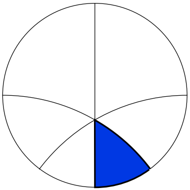
When $\mathcal{F}$ meets $\partial\mathbb{H}$ in an interval, a *funnel* occurs in $\Gamma\backslash\mathbb{H}$.

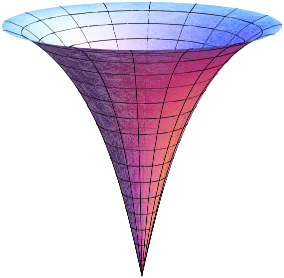




Conical points

An elliptic fixed point of Γ causes a conical singularity in $\Gamma\backslash\mathbb{H}$.





There is a 1-1 correspondence:

$$\text{conical points} \longleftrightarrow \text{orbits of elliptic fixed points.}$$

III. Spectral Theory

Influenced by the work of Maass, Selberg pioneered the study of the spectral theory of hyperbolic surfaces in the 1950's. The idea was to bring techniques from harmonic analysis into the study of automorphic forms.

Laplacian

The Laplacian operator on $\mathbb{H}$ is

$$\Delta := -y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right)$$

Since the Laplacian is invariant under isometries (for any metric),

$$\Delta \circ g = g \circ \Delta, \qquad \text{for } g \in \mathrm{PSL}(2, \mathbb{R}).$$

Hence the action of Δ on automorphic functions is well-defined.

The measure associated to the hyperbolic metric,

$$dA(z) = \frac{dx\,dy}{y^2},$$

is invariant under $\mathrm{PSL}(2,\mathbb{R})$ and thus defines a measure on $\Gamma\backslash\mathbb{H}$.

The resulting Hilbert space is $L^2(\Gamma\backslash\mathbb{H}, dA)$, with

$$\langle f, g\rangle = \int_{\mathcal{F}} \overline{f(z)}\, g(z)\, dA(z),$$

for any fundamental domain $\mathcal{F}$.

To define Δ as a self-adjoint operator acting on $L^2(\Gamma\backslash\mathbb{H}, dA)$, we apply the Friedrichs extension to Δ on the domain

$$\mathcal{D} := \Big\{ f \in C_0^\infty(\Gamma\backslash\mathbb{H}) : f \text{ and } \Delta f \in L^2(\Gamma\backslash\mathbb{H}, dA) \Big\}.$$

(Our sign convention is $\Delta \geq 0$.)

Eigenvalues

An *eigenvalue* of Δ on $\Gamma\backslash\mathbb{H}$ is λ such that

$$\Delta\phi = \lambda\phi,$$

where $\phi \in L^2(\Gamma\backslash\mathbb{H}, dA)$.

In the context of automorphic forms, eigenvectors of the Laplacian are called *Maass forms*.

If $\Gamma\backslash\mathbb{H}$ is compact, then the eigenvalues fill out the spectrum. There is an orthonormal basis of eigenfunctions $\{\phi_j\}$, and the corresponding eigenvalues satisfy

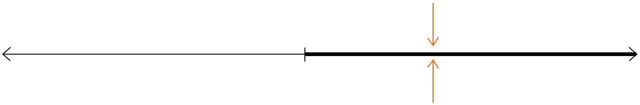
$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \to \infty.$$

Resolvent

If $\Gamma\backslash\mathbb{H}$ is non-compact, then we understand the spectrum by studying the resolvent $(\Delta - \lambda)^{-1}$

Stone's formula gives spectral projectors in terms of limits of as λ approaches $[0, \infty)$ from above and below.

$$P_{\alpha,\beta} = \lim_{\varepsilon \to 0} \frac{1}{2\pi i} \int_{\alpha}^{\beta} \left[(\Delta - \lambda - i\varepsilon)^{-1} - (\Delta - \lambda + i\varepsilon)^{-1} \right] d\lambda$$



In $\mathbb{H}$,

$$\Delta y^s = s(1-s)y^s,$$

suggesting that $\lambda = s(1-s)$ is a natural substitution for the spectral parameter.

Indeed, if we write the resolvent as

$$R_{\mathbb{H}}(s) := (\Delta - s(1-s))^{-1},$$

for $\operatorname{Re} s > \frac{1}{2}$, the corresponding integral kernel is

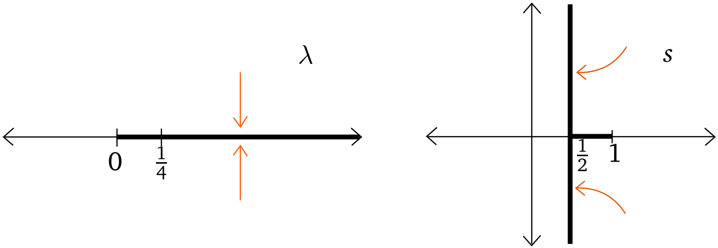
$$R_{\mathbb{H}}(s; z, w) = \frac{1}{4\pi} \frac{\Gamma(s)^2}{\Gamma(2s)} \cosh^{-s}(d/2)\, F\big(s, s; 2s; \cosh^{-1}(d/2)\big),$$

where $d := d(z, w)$.

Meromorphic continuation

The same picture holds for $\Gamma\backslash\mathbb{H}$: the resolvent $R(s) = (\Delta - s(1-s))^{-1}$ admits a meromorphic continuation to $s \in \mathbb{C}$.

The set $\lambda \in [0, \infty)$ corresponds to $s \in [\frac{1}{2}, 1] \cup \{\operatorname{Re} s = \frac{1}{2}\}$.



From the behavior of $R(s)$ as s approaches the critical line, we can understand the spectral projectors.

How do we continue the resolvent? By constructing a parametrix.

1. Use $R(s_0)$ for large $\operatorname{Re} s_0$ in the interior.
2. In cusps and funnels, use model resolvents for cylindrical quotients to construct $R^0(s)$.
3. Paste these together using cutoffs χ_j, to get the parametrix

$$M(s) := \chi_2 R(s_0)\chi_1 + (1-\chi_0)R^0(s)(1-\chi_1),$$

and compute the error

$$(\Delta - s(1-s))M(s) = I - K(s).$$

4. Show that $K(s)$ is compact on a weighted L^2 space, and use the Analytic Fredholm Theorem to invert

$$R(s) = M(s)(I-K(s))^{-1}.$$

Spectrum of Δ

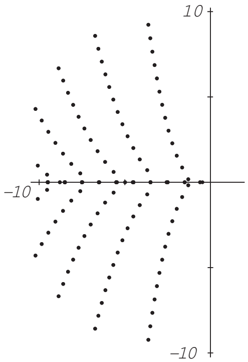
From the structure of the resolvent, we can deduce basic spectral information.

1. For $\Gamma\backslash\mathbb{H}$ compact, Δ has discrete spectrum in $[0, \infty)$.
2. For $\Gamma\backslash\mathbb{H}$ non-compact, Δ has absolutely continuous spectrum $[\frac{1}{4}, \infty)$ and discrete spectrum contained in $[0, \infty)$.
3. If $\Gamma\backslash\mathbb{H}$ has infinite-area, then there are no embedded eigenvalues, i.e. the discrete spectrum is finite and contained in $(0, \frac{1}{4})$.

Resonances

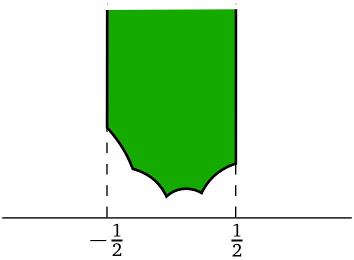
The poles of the meromorphically continued resolvent $R(s)$ are called *resonances*.

These include points with $\operatorname{Re} s \geq \frac{1}{2}$ for which $\lambda = s(1-s)$ is a discrete or embedded eigenvalues, and possibly other points with $\operatorname{Re} s < \frac{1}{2}$.



Cusp forms

For simplicity, assume that $\Gamma\backslash\mathbb{H}$ has finite area with a single cusp. We can assume that the cusp corresponds to the point $\infty \in \mathbb{H}$, with $\Gamma_\infty := \langle z \mapsto z+1 \rangle \subset \Gamma$ and $\mathcal{F}$ bounded by $|\operatorname{Re} s| = \frac{1}{2}$.



The *cusp forms* are defined by

$$\mathcal{H}_{\text{cusp}} := \left\{ f \in L^2(\mathcal{F}, dA) : \int_0^1 f\,dx = 0 \text{ for a.e. } y > 0 \right\}.$$

In other words, if we expand f in a Fourier series

$$f(z) = \sum c_n(y) e^{2\pi i n x},$$

the $n = 0$ coefficient vanishes for $y > 0$.

For solutions of $(\Delta - s(1-s))f = 0$ the coefficient $c_0(y)$ behaves like y^s or y^{1-s} as $y \to \infty$, while the non-zero modes are Bessel functions that either grow or decay exponentially in y.

This fact implies that the restriction of Δ to $\mathcal{H}_{\text{cusp}}$ has purely discrete spectrum.

Maass cusp forms

The eigenvectors of the restriction of Δ to $\mathcal{H}_{\text{cusp}}$ are called *Maass cusp forms*.

Embedded eigenvalues ($\lambda \geq \frac{1}{4}$) are automatically cusp forms. For $\lambda < \frac{1}{4}$, eigenvalues may or may not be cusp forms.

Selberg showed that certain arithemetic surfaces have an abundance of Maass cusp forms. But Phillips and Sarnak showed that these disappear when the arithmetic surface is deformed to an ordinary hyperbolic surface. They conjectured that $\mathcal{H}_{\text{cusp}}$ is small or empty for a general cofinite Fuchsian group.

Eisenstein series

(We continue to assume one cusp at ∞ with $\Gamma_\infty \subset \Gamma$.)

Since y^s solves the eigenvalue equation and is invariant under Γ_∞, we can try to solve the eigenvalue equation on $\Gamma\backslash\mathbb{H}$ by averaging

$$\begin{aligned} E(s;z) &:= \sum_{g\in\Gamma_\infty\backslash\Gamma} (\operatorname{Im} gz)^s \\ &= \sum_{g\in\Gamma_\infty\backslash\Gamma} \frac{y^s}{|cz+d|^{2s}}. \end{aligned}$$

This is called an *Eisenstein series*. It converges for $\operatorname{Re} s > 1$, but by connecting it to $R(s)$, we can show that it extends meromorphically to $s \in \mathbb{C}$.

Eisenstein series give a way to parametrize the continuous spectrum.

Spectral decomposition

The *residual spectrum* consists of eigenvalues with $\lambda < \frac{1}{4}$ which do not come from cusp forms. The span of such eigenvectors is denoted $\mathcal{H}_{\text{res}}$.

On $\mathcal{H}_{\text{cusp}} \oplus \mathcal{H}_{\text{res}}$ we have an complete eigenbasis $\{\phi_j, \lambda_j\}$ for Δ.

To complete our decomposition, define the continuous Hilbert space $\mathcal{H}_{\text{cont}} = (\mathcal{H}_{\text{cusp}} \oplus \mathcal{H}_{\text{res}})^{\perp}$, so that

$$L^2(\Gamma\backslash\mathbb{H}, dA) = \mathcal{H}_{\text{cusp}} \oplus \mathcal{H}_{\text{res}} \oplus \mathcal{H}_{\text{cont}}.$$

The spectrum of Δ on $\mathcal{H}_{\text{cont}}$ is purely continuous, and Eisenstein series give the spectral decomposition on this subspace.

This means that general $f \in L^2(\Gamma \backslash \mathbb{H}, dA)$ can be writen

$$f(z) = \int_{-\infty}^{\infty} E(\tfrac{1}{2} + ir; z) a(r)\, dr + \sum b_j \phi_j(z),$$

where

$$a(r) := \int_{\mathcal{F}} \overline{E(\tfrac{1}{2} + ir; z)} f(z)\, dA(z),$$

$$b_j := \langle \phi_j, f \rangle.$$

E(s; z) is the cusp analog of a plane wave.

Scattering matrix

As $y \to \infty$,

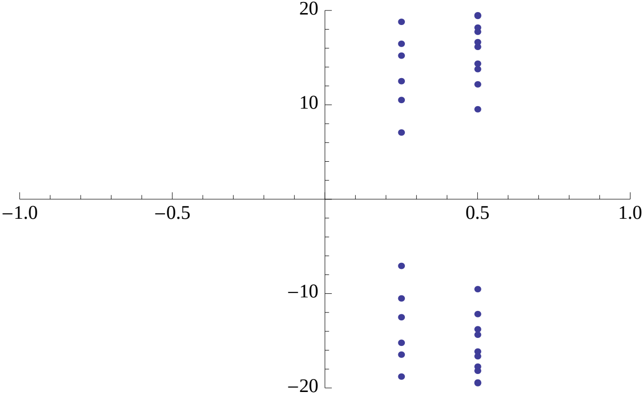
$$E(s; z) = y^s + \varphi(s) y^{1-s} + O(y^{-\infty}),$$

for a meromorphic function $\varphi(s)$ called the *scattering matrix* (or *scattering determinant*, in this one-cusp case).

The poles of $\varphi(s)$ are the *scattering poles*. These include points where $\lambda = s(1 - s)$ is in the residual spectrum.

Cusp are resonances but not scattering poles.

resonances = scattering poles + cusp resonances



Multiple cusps

If there are many cusps, for each cusp we consider a model where the corresponding parabolic fixed point is moved to ∞. The y coordinate for the j-th cusp is denoted y_j.

For cusp forms, the vanishing condition applies separately to each cusp.

For cusp i, the Eisenstein series $E_i(s; z)$ is defined as an average of y_i over Γ. We then consider the asymptotic expansion in the j-th cusp:

$$E_i(s; z) \sim \delta_{ij} y^s + \varphi_{ij}(s) y_j^{1-s}.$$

The scattering matrix is now an actual matrix $[\varphi_{ij}]$, with scattering determinant $\varphi(s)$.

Infinite-area case

The presence of a single funnel means there can be no cusp forms. In particular, the discrete spectrum is finite and contained in $(0, \frac{1}{4})$. There are no embedded eigenvalues.

We can still parametrize the continuous spectrum by the analog of Eisenstein series. These analogs are now called *Poisson operators*.

Asymptotic expansions of Poisson operators define the scattering matrix, for which the funnel components are pseudodifferential operators.

IV. Selberg trace formula

The central result in the spectral theory of hyperbolic surfaces is the Selberg trace formula.

To any $f \in C^\infty[0,\infty)$ we can try to define an operator K_f with integral kernel

$$K_f(z,w) := \sum_{g \in \Gamma} f(d(z,gw)).$$

(The sum will only converge if f decays sufficiently at ∞.)

The Selberg trace formula computes the trace of K_f in two different ways.

Traces in the compact case

Suppose $\Gamma\backslash\mathbb{H}$ is compact, with spectrum $\{\lambda_j\}$ and eigenvectors $\{\phi_j\}$.

Since K_f is a smoothing operator on $\Gamma\backslash\mathbb{H}$, the trace could be written as

$$\operatorname{tr} K_f = \int_{\Gamma\backslash\mathbb{H}} K_f(z,z)\, dA(z).$$

On the other hand, if the eigenvalues of K_f are $\{\kappa_j\}$, then

$$\operatorname{tr} K_f = \sum_j \kappa_j.$$

Spectral trace computation

Many geometric operators have the same form as K_f, such as the resolvent and heat operators.

K_f always commutes with Δ is diagonalized by the basis $\{\phi_j\}$.

To compute κ_j, consider the eigenvalue equation

$$\begin{aligned}\kappa_j\phi_j(w) &= \int_{\mathcal{F}} K_f(w,z)\,\phi_j(z)\,dA(z)\\ &= \sum_{g\in\Gamma}\int_{\mathcal{F}} f(d(w,gz))\,\phi_j(z)\,dA(z)\\ &= \int_{\mathbb{H}} f(d(w,z))\,\phi_j(z)\,dA(z)\end{aligned}$$

Now set $w = i$ and $s_j = \sqrt{\lambda_j - 1/4}$.

We can exploit the fact that y^{s_j} solves the same eigenvalue equation as ϕ_j,

$$(\Delta - \lambda_j)y^{s_j} = 0,$$

to prove

$$\kappa_j \phi_j(i) = \int_{\mathbb{H}} f(d(i,z))\, \phi_j(z)\, dA(z) = \phi_j(i) \int_{\mathbb{H}} f(d(w,z))\, y^{s_j}\, dA(z).$$

This gives κ_j in terms of f,

$$\kappa_j = \int_{\mathbb{H}} f(d(i,z))\, y^{s_j}\, dA(z)$$

Length trace computation

Assume now that $\Gamma\backslash\mathbb{H}$ is smooth and compact. This means that Γ contains only hyperbolic elements. Recall that closed geodesics of $\Gamma\backslash\mathbb{H}$ correspond to conjugacy classes of hyperbolic elements.

The length trace computation starts from

$$\begin{aligned} \operatorname{tr} K_f &= \int_{\mathcal{F}} K_f(z,z)\, dA(z) \\ &= \sum_{g\in\Gamma} \int_{\mathcal{F}} f(d(z,gz))\, dA(z). \end{aligned}$$

The trick is to organize the sum over Γ as a sum over conjugacy classes, and then express these in terms of lengths of closed geodesics.

Decomposition of Γ

Let Π be a complete list of inconjugate primitive elements of Γ. Then we can write

$$\Gamma - \{I\} = \bigcup_{g\in\Pi} \bigcup_{k\in\mathbb{N}} \bigcup_{h\in\Gamma/\langle g\rangle} \{hg^k h^{-1}\}.$$

Associated to each $g \in \Pi$ is a primitive closed geodesic of $\Gamma\backslash\mathbb{H}$, and the corresponding lengths form the primitive length spectrum:

$$\mathcal{L}(\Gamma) := \{\ell(g) : \ g \in \Pi\}.$$

(Note that g and g^{-1} are not conjugate; this is the ‘oriented’ length spectrum.)

Sum over Γ

Using the conjugacy class decomposition of Γ, we write the trace as

$$\operatorname{tr} K_f = f(0)\ \operatorname{Area}(\Gamma\backslash\mathbb{H}) + \sum_{g\in\Pi}\sum_{k\in\mathbb{N}}\sum_{h\in\Gamma/\langle g\rangle}\int_{\mathcal{F}} f(d(z, hg^k h^{-1}z))\, dA(z)$$

By a change of variables, we can write

$$\int_{\mathcal{F}} f(d(z, hg^k h^{-1}z))\, dA(z) = \int_{h\mathcal{F}} f(d(z, g^k z))\, dA(z)$$

The union of $h\mathcal{F}$ over $\Gamma/\langle g\rangle$ is a fundamental domain for the cyclic group $\langle g\rangle$. We could replace this by any other fundamental domain, so

$$\sum_{h\in\Gamma/\langle g\rangle}\int_{\mathcal{F}} f(d(z, hg^k h^{-1}z))\, dA(z) = \int_{\mathcal{F}_{\langle g\rangle}} f(d(z, g^k z))\, dA(z).$$

We can conjugate g to $z \mapsto e^{\ell} z$ and take $\mathcal{F}_{\langle g \rangle} = \{1 \leq y \leq e^{\ell}\}$, so that

$$\sum_{h \in \Gamma / \langle g \rangle} \int_{\mathcal{F}} f(d(z, hg^k h^{-1} z))\, dA(z) = \int_{\{1 \leq y \leq e^{\ell}\}} f(d(z, e^{k\ell} z))\, dA(z).$$

The integral evaluates to

$$\frac{\ell}{\sinh(k\ell/2)} \int_{k\ell}^{\infty} \frac{f(\cosh t)}{\sqrt{2\cosh t - 2\cosh k\ell}} \sinh t \, dt.$$

With this computation, we have

$$\begin{aligned}\operatorname{tr} K_f = & f(0)\ \operatorname{Area}(\Gamma\backslash\mathbb{H}) \\ & + \sum_{\ell\in\mathcal{L}(\Gamma)} \sum_{k\in\mathbb{N}} \frac{\ell}{\sinh(k\ell/2)} \int_{k\ell}^{\infty} \frac{f(\cosh t)}{\sqrt{2\cosh t - 2\cosh k\ell}} \sinh t\, dt.\end{aligned}$$

Recall that the spectral computation gave

$$\operatorname{tr} K_f = \sum_{j=0}^{\infty} h\Big(\sqrt{\lambda_j - \tfrac{1}{4}}\Big),$$

where

$$h(r) := \int_{\mathbb{H}} f(d(i,z))\, y^r\, dA(z).$$

Selberg trace formula (smooth compact case)

In terms of the function h,

$$\sum_{j=0}^{\infty} h\left(\sqrt{\lambda_j - \tfrac{1}{4}}\right) = \frac{\mathrm{Area}(\Gamma\backslash\mathbb{H})}{4\pi} \int_{-\infty}^{\infty} r h(r) \tanh \pi r \, dr + \sum_{\ell \in \mathcal{L}(\Gamma)} \sum_{k \in \mathbb{N}} \frac{\ell}{\sinh(k\ell/2)} \widehat{h}(k\ell).$$

General finite area case

The Trace Formula is somewhat trickier to prove for non-compact $\Gamma\backslash\mathbb{H}$, because the operator K_f is not trace class. And the integral

$$\int_{\Gamma\backslash\mathbb{H}} K_f(z,z)\, dA(z)$$

diverges!

To prove the trace formula, we must restrict operators to $y_j \leq N$ in each cusp and then carefully take the limit $N \to \infty$.

On the spectral side, extra ‘scattering’ terms appear to account for the continuous spectrum:

$$\sum_j h\left(\sqrt{\lambda_j - \tfrac{1}{4}}\right) - \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\varphi'}{\varphi}(\tfrac{1}{2} + ir)h(r)\,dr + \tfrac{1}{2}h(0)\,\mathrm{tr}[\varphi_{ij}(\tfrac{1}{2})],$$

where $[\varphi_{ij}(s)]$ is the scattering matrix and $\varphi(s)$ the scattering determinant.

On the length side, our decomposition of Γ must include conjugacy classes of primitive elliptic and parabolic elements, as well as hyperbolic.

For each elliptic fixed point of order m we pick up a term

$$\sum_{k=1}^{m-1} \frac{1}{m\sin(\pi k/m)} \int_{-\infty}^{\infty} \frac{e^{-2\pi kr/m}}{1-e^{-2\pi r}}\, h(r)\, dr.$$

For each cusp, we add a term

$$-\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\Gamma'}{\Gamma}(1+ir)h(r)\, dr + \tfrac{1}{2}h(0) - \check{h}(0)\log 4$$

Selberg zeta function

Many applications of the theory come through the Selberg zeta function,

$$Z(s) := \prod_{\ell \in \mathcal{L}(\Gamma)} \prod_{k=1}^{\infty} \left(1 - e^{-(s+k)\ell}\right).$$

Roughly speaking, the logarithmic derivative of $Z(s)$ is what appears on the length side when we take the trace of the resolvent.

The product converges for $\operatorname{Re} s \geq 1$, but spectral methods can be used to prove that $Z(s)$ has a meromorphic extension.

One nice feature of $Z(s)$ is that its zeros are essentially the resonances. (There are some extra 'topological' poles and zeros.)

Applications of the trace formula

For the finite-area case:

1. The resonance set and the length spectrum determine each other, and also χ and number of cusps.

2. Prime Geodesic Theorem:

$$\#\{e^{\ell} \leq x\} \sim \operatorname{Li} x + \sum \operatorname{Li}(x^{s_j}),$$

   where $\{s_j(1-s_j)\}$ are the eigenvalues in $(0, \frac{1}{4})$.

3. Weyl-Selberg asymptotic formula

$$\#\{|\lambda_j| \leq r\} - \frac{1}{4\pi}\int_{-\sqrt{r-1/4}}^{\sqrt{r-1/4}} \frac{\varphi'}{\varphi}(\tfrac{1}{2}+it)\,dt \sim \frac{\operatorname{Area}(\Gamma\backslash\mathbb{H})}{4\pi}\,r$$

V. Arithmetic surfaces

The term *arithmetic* implies a restriction to integers.

Consider a finite dimensional representation $\rho : \mathrm{PSL}(2,\mathbb{R}) \to \mathrm{GL}(n,\mathbb{R})$. Restricting to integer entries gives a Fuchsian group,

$$\Gamma := \left\{ g \in \mathrm{PSL}(2,\mathbb{R}) : \ \rho(g) \in \mathrm{GL}(n,\mathbb{Z}) \right\}.$$

Such groups, along with their subgroups of finite index, are called *arithmetic Fuchsian groups*.

Modular group

The prototype of an arithmetic Fuchsian group is the modular group

$$\Gamma_{\mathbb{Z}} := \mathrm{PSL}(2,\mathbb{Z}).$$

The group is generated by the elements

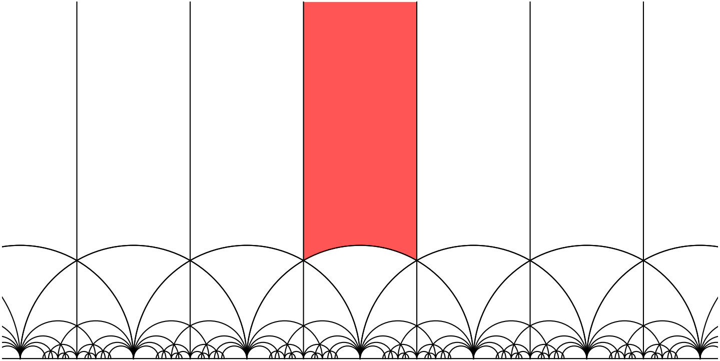
$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

or

$$T : z \mapsto z+1, \qquad S : z \mapsto -\frac{1}{z}.$$

The standard fundamental domain for $\Gamma_{\mathbb{Z}}$ is

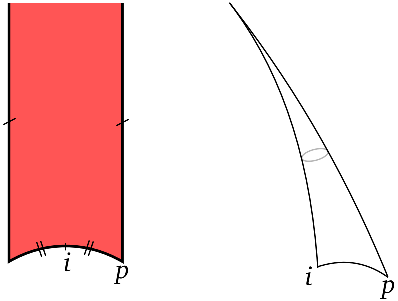
$$\mathcal{F} = \left\{ z \in \mathbb{H} : \; |\operatorname{Re} z| \leq \tfrac{1}{2}, |z| \geq 1 \right\}$$



Modular surface

The *modular surface* is $X := \Gamma_{\mathbb{Z}} \backslash \mathbb{H}$.

The generator S fixes i, and the point $p = e^{i\pi/3}$ is fixed by $ST^{-1}p$.



Gauss-Bonnet gives $\text{Area}(X) = \frac{\pi}{3}$.

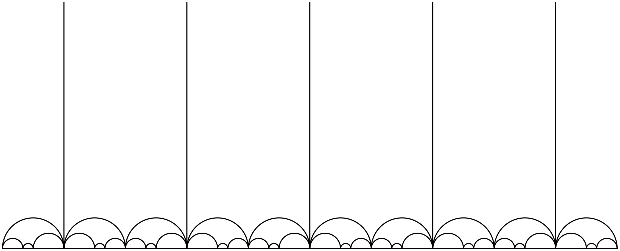
Congruence subgroups

For $N \geq 1$ the *principal congruence subgroup* of level N is

$$\Gamma_{\mathbb{Z}}(N) := \{g \in \Gamma_{\mathbb{Z}} : \ g \equiv I \quad \text{mod } N\}$$

For example

$$\Gamma_{\mathbb{Z}}(2) = \left\{ \begin{pmatrix} \text{odd} & \text{even} \\ \text{even} & \text{odd} \end{pmatrix} \subset \mathrm{PSL}(2, \mathbb{Z}) \right\}$$



The quotients are the modular surfaces,

$$X(N) := \Gamma_{\mathbb{Z}}(N)\backslash\mathbb{H}.$$

The surface $X(2)$ is a sphere with 3 cusps and area 2π.

The geometry gets more complicated as N increases - the genus of $X(N)$ is approximately N^3 for N large.

Number theorists are particularly interested in Maass cusp forms on $X(N)$.

Spectral theory of the modular surface

Let continue with $\Gamma_{\mathbb{Z}}$ and the modular surface X.

Recall that the scattering matrix was defined in terms of the Eisenstein series

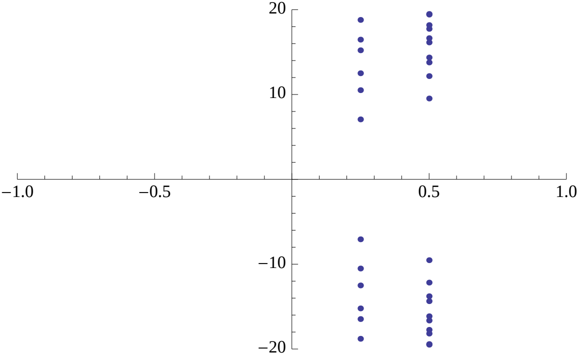
$$E(s;z) = \sum_{\Gamma_\infty \backslash \Gamma_{\mathbb{Z}}} \frac{y^s}{|cz+d|^{2s}}.$$

In this case, we can compute the asymptotic expansion explicitly, giving the scattering matrix

$$\varphi(s) = \sqrt{\pi}\, \frac{\Gamma(s-\frac{1}{2})}{\Gamma(s)} \frac{\zeta(2s-1)}{\zeta(2s)},$$

where $\zeta(z)$ is the Riemann zeta function. Meromorphic continuation is thus clear, and the scattering poles are solutions of $\zeta(2s) = 0$.

The modular surface also has many cusp forms:



Cusp eigenvalue $\lambda_j = \frac{1}{4} + r_j^2 \longrightarrow$ resonances at $s = \frac{1}{2} \pm i r_j$.

Riemann zeros $\longrightarrow$ scattering poles on $\mathrm{Re}\, s = \frac{1}{4}$.

The Selberg trace formula implies that

$$\#\left\{\text{cusp resonances with } |s - \tfrac{1}{2}| \le t\right\} \sim \frac{t^2}{6}.$$

In contrast, the asymptotics of the Riemann zeros gives

$$\#\left\{\text{scattering poles with } |s - \tfrac{1}{2}| \le t\right\} \sim \frac{2t \log t}{\pi}.$$

The cusp forms dominate the spectrum.

Hecke operators

Why makes arithmetic surfaces so special?

They have hidden symmetries, called *Hecke operators*.

For $f \in L^2(\Gamma_{\mathbb{Z}} \backslash \mathbb{H}, dA)$ and $n \in \mathbb{N}$, define

$$T_n f(z) := \frac{1}{\sqrt{n}} \sum_{ad=n} \sum_{b=0}^{d-1} f\left(\frac{az+b}{d}\right).$$

It's clear that T_n commutes with Δ. What is not so obvious is that $T_n f$ is still invariant under $\Gamma_{\mathbb{Z}}$.

As an example, consider

$$T_3 f(z) = \frac{1}{\sqrt{3}}\left[f(3z) + f\left(\frac{z}{3}\right) + f\left(\frac{z+1}{3}\right) + f\left(\frac{z+2}{3}\right)\right].$$

Invariance under $z \mapsto z+1$ is actually obvious.

The other generator is $z \mapsto -1/z$, so consider

$$T_3 f\left(-\frac{1}{z}\right) = \frac{1}{\sqrt{3}}\left[f\left(-\frac{3}{z}\right) + f\left(-\frac{1}{3z}\right) + f\left(\frac{z-1}{3z}\right) + f\left(\frac{2z-1}{3z}\right)\right].$$

The first 2 terms just switched places:

$$f\left(-\frac{3}{z}\right) = f\left(\frac{z}{3}\right), \quad f\left(-\frac{1}{3z}\right) = f(3z).$$

For the other terms we can use the matrix identity,

$$\begin{pmatrix} 1 & -1 \\ 3 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 3 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$$

(note that $\det \begin{pmatrix} 1 & -1 \\ 3 & -2 \end{pmatrix} = 1$!), to see that

$$f\left(\frac{z-1}{3z}\right) = f\left(\frac{z+2}{3}\right).$$

Similarly,

$$f\left(\frac{2z-1}{3z}\right) = f\left(\frac{z+1}{3}\right).$$

Why does this work?

Let

$$\Upsilon_n := \{A \in GL(2,\mathbb{Z}) : \det A = n\},$$

and define an equivalence relation on Υ_n,

$$A_1 \sim A_2 \quad \text{if} \quad A_1 \in \Gamma_{\mathbb{Z}} A_2.$$

List the matrices occurring in the definition of T_n:

$$\mathcal{A}_n = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} : ad = n,\ 0 \le b \le d-1 \right\}.$$

Claim: The partition of Υ_n into distinct equivalence classes is

$$\Upsilon_n = \bigcup_{A \in \mathcal{A}_n} [A]$$

Consider

$$T_n f(z) = \frac{1}{\sqrt{n}} \sum_{A \in \mathcal{A}_n} f(Az).$$

Suppose $g \in \Gamma_{\mathbb{Z}}$ and $A \in \mathcal{A}_n$. Then $\det(Ag) = n$, so $Ag \in \Upsilon_n$. Our claim then says that

$$Ag = hA',$$

for $h \in \Gamma_{\mathbb{Z}}$ and $A' \in \mathcal{A}_n$.

If f is automorphic,

$$f(Agz) = f(hA'z) = f(A'z),$$

so

$$T_n f(gz) = \frac{1}{\sqrt{n}} \sum_{A' \in \mathcal{A}_n} f(A'z) = T_n f(z).$$

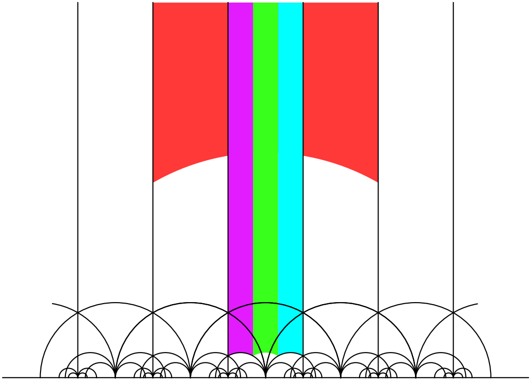
The full ring of Hecke operators T_n together with Δ can be simultaneously diagonalized.

In particular, we can choose a basis of Maass cusp forms such that

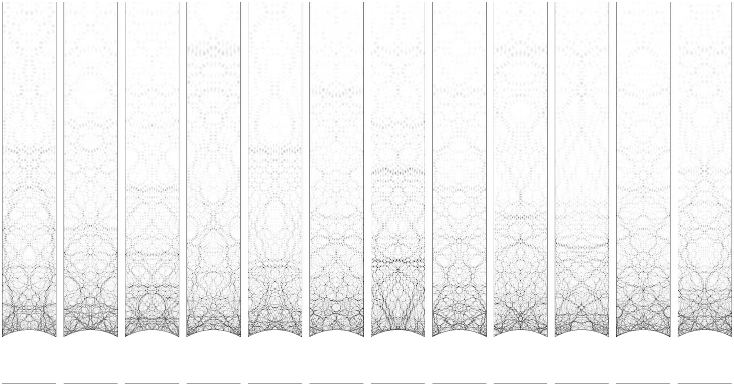
$$\Delta\phi_j = \lambda_j\phi_j, \qquad T_n\phi_j = \tau_j(n)\phi_j.$$

Indeed, it is conjectured that the cusp spectrum of X is simple, which would imply that Maass cusp forms are automatically Hecke eigenfunctions.

What do the Hecke symmetries look like? Here's a picture for T_3.



Compare to density plots of Maass cusp forms:



(Alex Barnett/Holger Then)

L-function connection

One primary reason that Maass cusp forms are important in number theory is the connection to L-functions (series which generalize the Riemann zeta function).

If ϕ is a Maass cusp form, with Hecke eigenvalues $\tau(n)$, then we can form a related L-series

$$L(s,\phi) := \sum_n \tau(n) n^{-s} = \prod_p \frac{1}{1-\tau(p)p^{-s}-p^{-2s}}.$$

The L function converges for $\operatorname{Re} s$ large, and the modularity of ϕ is equivalent to the analytic continuation of $L(s,\phi)$, with a functional relation connecting the values at s and $1-s$.

Many spectral questions about Maass cusp forms are related to properties of L-functions.

Artin's conjecture connects two-dimensional representations of Galois groups, via the associated L-functions, to Maass cusp forms with eigenvalue $\frac{1}{4}$.

On a related note, the Selberg-Ramanujan conjecture says that $X(N)$ has no discrete spectrum below $\frac{1}{4}$.

Quantum ergodicity

On the high-frequency side one of the main questions has been the asymptotic distribution of cusp forms.

The geodesic flow of a hyperbolic surface is ergodic, meaning that the only functions invariant under the geodesic flow are constant a.e.

Classical ergodicity is expected to correspond with 'quantum ergodicity', meaning equidistribution of eigenvectors as $j \to \infty$.

Quantum ergodicity is interpreted in terms of probability measures associated to eigenfunctions:

$$\phi_j \mapsto \nu_{\phi_j} = |\phi_j|^2 \, dA.$$

(There is a related phase space version, μ_{ϕ_j} on the unit tangent bundle, called the microlocal lift).

The conjecture is that ν_{ϕ_j} approaches dA in some sense (or μ_{ϕ_j} approaches Liouville measure on the unit tangent bundle).

For compact manifolds with ergodic geodesic flow, this is known to be true for ‘most' subsequences $j_k \to \infty$. Quantum *unique* ergodicity is the question of whether it holds for all sequences.

For arithmetic surfaces, QUE becomes a problem of estimating certain special values of L-functions, and it is this connection that has lead to a proof in this case.

General references:

1. S. Katok, *Fuchsian Groups*, University of Chicago Press, Chicago, IL, 1992.
2. P. Buser, *Geometry and Spectra of Compact Riemann Surfaces*, Birkhäuser Boston Inc., Boston, MA, 1992.
3. A.B. Venkov, *Spectral Theory of Automorphic Functions and Its Applications*, Kluwer Academic Publishers, Dordrecht, 1990.
4. D. Borthwick, *Spectral Theory of Infinite-Area Hyperbolic Surfaces*, Birkhäuser Boston Inc., Boston, MA, 2007.
5. T.M. Apostol, Modular functions and Dirichlet series in number theory. Springer-Verlag, NY, 1990
6. P. Sarnak, Spectra of hyperbolic surfaces, lecture notes, 2003.
7. P. Sarnak, Recent progress on QUE, lecture notes, 2009.