SPECTRAL THEORY OF AUTOMORPHIC FORMS

LECTURES BY PETER SARNAK,
NOTES BY TONY FENG AND GYUJIN OH

CONTENTS

SYLLABUS

Outline of the course. This is a course on the spectral theory of automorphic forms. The main goal is to explain the meromorphic continuation of Eisenstein series, due to Selberg and Langlands. We will follow the 1979 notes of Cohen-Sarnak (available on my website). This explains the proof of Selberg (1966), which was rediscovered by Bernstein in the 80s.

In the first half, we will be working on the upper half plane $\mathbf{H}^n$. A reference is my 2004 paper [Sar03] in the Bulletin of the AMS. In the second half, we'll study the higher rank case.

Prerequisites. We will assume background in the following subjects:

(1) Basic real, complex, and functional analysis (Fredholm theory).
(2) Modular forms.
(3) Representation theory of compact groups.
(4) Basic Riemannian geometry.

1. OVERVIEW

I am going to give an overview of the course: I'll explain the basic objects we'll be studying and some applications and motivations.

1.1. Spaces of constant curvature.

Let X be a Riemmanian manifold, with metric

$$ds^2 = \sum g_{ij} dx^i dx^j.$$

This induces a volume form

$$dx = \sqrt{g}\, dx_1 \dots dx_n,$$

and a Laplacian (using Einstein summation)

$$\Delta := \mathrm{div} \circ \mathrm{grad} = \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^i} \left(\sqrt{|g|} g^{ij} \frac{\partial}{\partial x^j} \right).$$

Example 1.1.1. Consider $\mathbf{R}^n$ with the usual metric, which has curvature 0. The isometry group $\mathrm{Isom}(\mathbf{R}^n)$ is generated by:

- translations $x \mapsto x + v$, and
- rotations.

We are interested in the decomposition of functions under operators commuting with the translations.[1] In $\mathbf{R}^n$ these are *convolution* operators, which are diagonalized by characters $e(\langle x, \xi \rangle)$. In general, we are looking for the analogues of these characters for a general semisimple group G.

Example 1.1.2. Consider S^n with the round metric, which has constant curvature 1. Then $\mathrm{Isom}(S^n)$ is the orthogonal group $O(n)$, and the irreducible representations in the space of functions are the spherical harmonics.

Example 1.1.3. Let $\mathbf{H}^{n+1}$ be $n+1$-dimensional hyperbolic space, which has constant curvature -1. A model for $\mathbf{H}^{n+1}$ is $\{(x, y) \colon x \in \mathbf{R}^n, y > 0\}$[2], with metric

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

The isometry group is $\mathrm{Isom}(\mathbf{H}^{n+1})$, which is generated by the following types of transformations:

(1) *Hyperbolic*:[3] $(y, x) \mapsto (\lambda y, \lambda \rho x)$, for $\lambda > 0$ and $\rho \in O(n)$. These fix the geodesics $\{(y, 0) \colon y > 0\}$.
(2) *Parabolic*: $(y, x) \mapsto (y, x + v)$ for $v \in \mathbf{R}^n$. These fix ∞.
(3) *Inversion*: $(y, x) \mapsto (\frac{y}{x^2+y^2}, \frac{x}{x^2+y^2})$.

Remark 1.1.4. For $n = 1$ the isometry group is $\mathrm{SL}_2(\mathbf{R})$, and these different types of transformations can be described in terms of the eigenvalues.

1.2. **Lattices.**

So far this has just been analysis; no number theory is present yet. The number theory enters when we introduce a *discrete group* Γ, which should be a "lattice" in the ambient group, and which will usually be defined through integers in some number field.

Example 1.2.1. Let Γ be a lattice in $\mathbf{R}^n$, i.e. Γ is a discrete subgroup with $\mathbf{R}^n/\Gamma$ compact. Hence $\Gamma \cong \mathbf{Z}^n$, and $\mathbf{R}^n/\Gamma$ is an n-torus $\mathbf{T}^m$. The eigenfunctions of Δ are $e(\langle x, \xi \rangle)$ for $\xi \in \Gamma^\vee \cong \mathbf{Z}^n$.

By convention we declare the "eigenvalues" to be those λ such that $\Delta \phi + \lambda \phi = 0$; with this convention the eigenvalues are $\lambda = 4\pi^2 |\xi|^2$ for $\xi \in \mathbf{Z}^n$.

---

[1] What if we ask for operators commuting with the full isometry group $\mathrm{Isom}(\mathbf{R}^n)$? The translation-invariant operators are the differential operators. The differential operators that commute with all the rotations are just polynomials in the Laplacian.

[2] This is called the *Poincaré model*. There are other models of hyperbolic space, and you want to choose the one best suited for your problem. In the theory of Eisenstein series the boundary is the most important aspect, so we choose a model that presents it well.

[3] In the jargon of the trace formula, these are called "elliptic"

1.3. **Poisson summation.** Let $f \in \mathscr{S}(\mathbf{R}^n)$ be a Schwartz function. Then $\widehat{f} \in \mathscr{S}(\mathbf{R}^n)$, where

$$\widehat{f}(\xi) = \int_{\mathbf{R}^n} e(\langle -x, \xi \rangle) f(x) dx.$$

The Poisson summation formula says that

$$\sum_{\nu \in \mathbf{Z}^\nu} f(\nu) = \sum_{m \in \mathbf{Z}^n} \widehat{f}(m). \tag{1.3.1}$$

We are going to give a proof of this, which is different from the usual one.

Proof. Let's define a kernel function on $L^2(\mathbf{R}^n/\mathbf{Z}^n)$ by

$$K(x, y) := \sum_{\nu \in \mathbf{Z}^n} f(x - y + \nu).$$

This certainly converges, as f is Schwartz, and evidently depends only on $x - y$. So we set $K(x - y) = K(x, y)$. We can then define an operator

$$K : L^2(\mathbf{R}^n/\mathbf{Z}^n) \to L^2(\mathbf{R}^n/\mathbf{Z}^n)$$

by

$$K \cdot h(x) = \int_{\mathbf{R}^n/\mathbf{Z}^n} K(x, y) h(y)\, dy.$$

Then K is Hilbert-Schmidt, which implies that all the nice facts about finite-dimensional linear algebra carry over in some form.[4] In particular, the trace is a "sum over the diagonal"

$$\operatorname{Tr} K = \int_{\mathbf{R}^n/\mathbf{Z}^n} K(x, x)\, dx,$$

which is manifestly equal to $\sum_\nu f(\nu)$.

Now the point is that we can compute the trace using a different basis. Let's choose the basis that diagonalizes the convolution operator, whose elements are

$$\phi_m(x) = e(-\langle x, m \rangle), \quad m \in \mathbf{Z}^n.$$

Let's compute the eigenvalue:

$$\begin{aligned} \int_{\mathbf{R}^n/\mathbf{Z}^n} K(x - y)\phi_m(y)\, dy &= \int_{\mathbf{R}^n/\mathbf{Z}^n} \sum_{\nu \in \mathbf{Z}^n} f(x - y + \nu) e(-\langle y, m \rangle)\, dy \\ &= \int_{\mathbf{R}^n} f(x - y) e(-\langle y, m \rangle)\, dy \\ &= \widehat{f}(n)\phi_m(x). \end{aligned}$$

The fact that trace of a diagonal matrix is just the sum of the eigenvalues then gives (1.3.1). □

Example 1.3.1. Consider $e(x^2 z)$ for $z \in \mathbf{H}$. If we sum over the integers, then we get a sum of the Fourier transform of the dual lattice. Since the Fourier transform of a Gaussian is a Gaussian, this shows that

$$\theta(z) := \sum_{n \in \mathbf{Z}} e(n^2 z)$$

[4] "The rule of the game is that if it makes sense, it's true".

is a modular form for $\Gamma_0(4) \subset \mathrm{SL}_2(\mathbf{Z})$. See Serre's book [Se73] for all this. Riemann used this to prove the analytic continuation and functional equation for $\zeta(s)$.[5]

1.4. **General setup.** We will consider a group G (which will be interpreted as the isometry group of some space), and a lattice $\Gamma \subset G$, which means that $\mathrm{Vol}(\Gamma\backslash G) < \infty$. We are mainly interested in the case where G is a semi-simple Lie group. Note that G is unimodular, so it does not matter if we consider the left or right Haar measure.

Example 1.4.1. Let $G = \mathrm{SL}_n(\mathbf{R})$, $K = \mathrm{SO}_n(\mathbf{R})$ the maximal compact subgroup. We'll be interested in the "symmetric space" $S = G/K$, which is the upper half plane, and in the "locally symmetric space" $X = \Gamma\backslash S$.

From a representation-theoretic point of view, we are interested in the space

$$L^2(\Gamma\backslash G) = \left\{ f : G \to \mathbf{C} \mid f(\gamma g) = f(g) \text{ for all } \gamma \in \Gamma; \int_{\Gamma\backslash G} |f(x)|^2 \, dg < \infty \right\}.$$

Remark 1.4.2. Why do we study L^2 instead of other function spaces? In fact the Eisenstein series won't be in L^2. But we want to work in a Hilbert space.

There is an obvious representation of G on $L^2(\Gamma\backslash G)$, by translation:

$$R_g f(x) = f(xg).$$

This defines an operator

$$R_g : L^2(\Gamma\backslash G) \to L^2(\Gamma\backslash G).$$

Moreover, it is an isometry, so we have a *unitary* representation of G.

We want to decompose the Hilbert space $L^2(\Gamma\backslash G)$ into irreducibles. Ideally this decomposition would have the shape

$$L^2(\Gamma\backslash G) = \bigoplus_{\pi \in \widehat{G}_{\mathrm{unit}}} m_\Gamma(\pi) H_\pi.$$

If $\Gamma\backslash G$ is compact then we do have such a decomposition. However, in the non-compact case there is also a continuous part, which is the Eisenstein series:

$$L^2(\Gamma\backslash G) = \bigoplus_{\pi \in \widehat{G}_{\mathrm{unit}}} m_\Gamma(\pi) H_\pi \oplus \int (\text{Eisenstein series}). \tag{1.4.1}$$

The main motivation for Selberg's meromorphic continuation of Eisenstein series was to make sense of the decomposition (1.4.1), and apply the trace formula.

1.5. **Matsushima's formula.** In the compact case, Matsushima's formula expresses the dimensions of $(\mathfrak{g}, K)$-cohomology groups in terms of the multiplicities $m_\Gamma(\pi)$.

Remark 1.5.1. This spectral decomposition can be used to prove vanishing of certain Betti numbers when the set of π's that contribute to Matsushima's formula is empty. This happens that when a representation that "wants" to contribute fails to be *unitary*. For examples, this happens for all finite index subgroups of $\mathrm{SL}_3(\mathbf{Z})$.

[5]This proof actually precedes Riemann; Weil discovered it as an exercise in a textbook (cf. [Wei87]).

1.6. **Homogeneous dynamics.** The subject of homogeneous dynamics studies orbits on spaces like $\Gamma\backslash G$. For SL_2, an orbit of the form

$$\Gamma x \begin{pmatrix} e^{t/2} & \\ & e^{-t/2} \end{pmatrix}$$

is called the *geodesic flow*. An orbit of the form

$$\Gamma x \begin{pmatrix} 1 & \\ t & 1 \end{pmatrix}$$

is called a *horocycle flow*. The point is that $\mathrm{SL}_2(\mathbf{R})$ is the unit sphere bundle (of the tangent bundle) over $\mathbf{H}$. Spectral theory can be used to understand things about these orbits, for e example whether or not they are equidistributed. However, one of the major theorems is *Ratner's theorem*, which lies deeper than spectral theory can access.

1.7. **The Ramanujan conjecture.** The only representation I really understand is the trivial representation. In some sense spectral theory is about showing that the trivial representation is the only one that matter - one wants to show that the other representations contribute negligibly. The *Ramanujan conjecture* predicts that all the other representations are very far from the trivial representation; more precisely, they are *tempered*. The thing which is important to me is to know that the other representations are tempered.

1.8. L**-functions.** For me the real interest in this subject comes from L-functions.[6]

Example 1.8.1. Let $G = \mathrm{GL}_n(\mathbf{R})$ and $\Gamma \le \mathrm{SL}_n(\mathbf{Z})$ be a finite index subgroup. We will prove later that $\mathrm{vol}(\Gamma\backslash G) < \infty$. We define a *congruence subgroup* $\Gamma(N) \subset \Gamma$ by

$$\Gamma(N) = \left\{\gamma \in \Gamma \colon \gamma \equiv \mathrm{Id} \pmod N\right\}.$$

We're interested in the part of the spectrum which is not Eisenstein, which is called *cuspidal.* Langlands was the first to appreciate it. We'll actually use Eisenstein series to deduce information about the cuspidal representations.

If π is a "cusp form" (cuspidal representation) then there is a "standard L-function" $L(s,\pi)$, which has all the properties (functional equation, meromorphic continuation) of $\zeta(s)$. This follows from Riemann's method, and was established by Tamagawa and Godement-Jacquet. This is *the* way to "grow" L-functions: all L-functions which have nice properties - meromorphic continuation, functional equation, etc. - are supposed to be obtainable in this way.

Consider the moduli space of lattices L in $\mathbf{R}^n$, with volume 1, is $\mathrm{SL}_n(\mathbf{Z})\backslash \mathrm{SL}_n(\mathbf{R})$. For $G = \mathrm{SL}_2(\mathbf{R})$, how can we make a lattice $\Gamma < \mathrm{SL}_2(\mathbf{R})$?

Let S be a Riemann surface of genus $g \ge 2$. By the Uniformization Theorem, the universal cover of S is $\mathbf{H}$ as a complex manifold. So there is a cocompact $\Gamma \le \mathrm{SL}_2(\mathbf{R})$ such that $S \cong \Gamma\backslash\mathbf{H}$. The moduli space of genus g Riemann surfaces is a complex manifold of dimension $3g-3$, if $g \ge 2$. Hence the set of such Γ's form *continuous families.* In particular, there are uncountably many Γ's, so they cannot all come from number theory. It is a very special feature of SL_2 that this happens!

[6] The theory of modular forms is not interesting in the function field world. There, Grothendieck's algebraic geometry gives all the understanding that you would need.

2. Symmetric spaces

2.1. **Basic definitions.** We will discuss the "soft theory" of symmetric spaces, following Selbert's 1956 paper [Sel56]. The "softness" comes from the fact that we will not compute anything.

Definition 2.1.1. A *symmetric space* is a connected complete Riemannian manifold for which the geodesic inversion in any point p is a global isometry. We fix notation for the Riemannian metric:

$$ds^2 = \sum_{i,j} g_{ij}\, dx^i\, dx^j.$$

Example 2.1.2. The symmetric spaces with constant curvature are $S^n, \mathbf{R}^n, \mathbf{H}^n$.

Example 2.1.3. Let $G = \mathrm{SL}_n(\mathbf{R})$, and

$$\mathscr{Y} := \{Y \in \mathrm{Mat}_{n\times n}(\mathbf{R})\colon Y \text{ positive definite}; \det Y = 1\}$$

with the metric

$$ds^2 = \mathrm{Tr}(Y^{-1}dY \cdot Y^{-1}dY)$$

This makes $\mathscr{Y}$ into a symmetric space, with the inverse $Y \mapsto Y^{-1}$ at Id being an isometry. G acts on $\mathscr{Y}$ isometrically by $y \mapsto g^t y g$.

Lemma 2.1.4. *For a symmetric space S, let $G := \mathrm{Isom}(S)$. Then G acts transitively on S.*

Proof. For any two points $x, y \in S$, we want to find $g \in G$ taking x to y. Take a geodesic between x and y and let g be the inversion about the midpoint. □

Let $K_{x_0} = \mathrm{Stab}(x_0)$ for $x_0 \in S$. Then by Lemma 2.1.4, we have $G/K \cong S$.

Remark 2.1.5. There is a classification of "irreducible" symmetric spaces, due to Cartan, in terms of presentations of the form G/K. See Helgason's book [H08].

If at x_0 we choose coordinates so that $g_{ij}(x_0) = \delta_{ij}$, then K_{x_0} acts on tangent vectors in a length-preserving manner, giving an embedding $K_{x_0} \hookrightarrow O(n)$. In particular, K_{x_0} is compact.

2.2. **Invariant differential operators.**

Definition 2.2.1. Let S be a symmetric space. For $g \in G$ define the operator

$$R_g f(x) := f(gx)$$

Let $\mathscr{D}(S)$ be the ring of linear invariant differential operators on S, i.e. a differential operator D lies in $\mathscr{D}(S)$ if and only if

$$R_g D = D R_g \text{ for all } g \in G.$$

Example 2.2.2. In $\mathbf{R}^n$ the condition forces constant coefficients.

To study the ring $\mathscr{D}(S)$ we will introduce the idea of point pair invariants.

Definition 2.2.3. A smooth function $k\colon S \times S \to \mathbf{C}$ is *point-pair invariant* if

$$k(\sigma x, \sigma y) = k(x, y) \text{ for all } \sigma \in G.$$

We also demand that k is continuous and compactly supported in the variable y for a given x.[7] We denote the algebra of point-pair invariants by $A(S)$.

Remark 2.2.4. In this formulation, the p-adic story, with symmetric spaces replaced by Bruhat-Tits buildings, is almost identical.

[7] However, sometimes it is useful to merely ask that k decays rapidly at infinity.

Think of these as kernels for the operators

$$k_1 \circ k_2(x,y) = \int_S k_1(x,w)k_2(w,y)\,dw.$$

This turns the point pair invariants into a ring.

Lemma 2.2.5. *The algebra $A(S)$ is commutative.*

Proof. Let μ be an inversion swapping $x \longleftrightarrow y$. Then we have

$$k(x,y) = k(\mu x, \mu y) = k(y,x).$$

So

$$\begin{aligned} k_1 \circ k_2(z,t) &= \int_S k_1(z,w)k_2(w,t)\,dw \\ &= \int k_1(\mu w, \mu z)k_2(\mu t, \mu w)\,dw \\ &= \int_S k_2(\mu t, w)k_1(w, \mu z)dw \\ &= k_2 \circ k_1(\mu t, \mu z) \\ &= k_2 \circ k_1(z,t) \end{aligned}$$

□

Proposition 2.2.6. *The ring $\mathscr{D}(S)$ is commutative.*

Proof. Let $\phi_\delta(x,y)$ be a point pair which is an approximation to the identity, i.e. $\phi_\delta(x,y)$ has compact support in y for fixed x, such that for smooth f

$$\int \phi_\delta(x,y)f(y) \to f(x) \text{ as } \delta \to 0.$$

We can arrange that $\phi_\delta(x,y)$ is a function of $d(x,y)$, which is supported in $[-\delta,\delta]$.

If $D \in \mathscr{D}(S)$, we'll show that it commutes with all point pair invariants, and then with each other. Let $k(x,y)$ be a smooth point pair invariant. Then you can check that $D_x k(x,y)$ is a point pair invariant invariant; here $D_x k$ means applying D to the first argument. (The point is that applying an invariant differential operator to a point pair invariant produces a point pair invariant.)

Hence $k \circ (D\phi_{\delta_1}(x,y)) = (D\phi_{\delta_1}(x,y)) \circ k$ by Lemma 1. Letting $\delta_1 \to 0$, we conclude that $k \circ D = D \circ k$ for all point pair invariants k. [This observation is not actually needed for the proof.]

Similarly, for $D_1, D_2 \in \mathscr{D}(S)$ we have

$$(D_1\phi_{\delta_1}) \circ (D_2\phi_{\delta_2}) = (D_2\phi_{\delta_2}) \circ (D_1\phi_{\delta_1})$$

hence taking $\delta_1, \delta_2 \to 0$ we conclude $D_1D_2 = D_2D_1$.

□

Let $\Delta = \operatorname{div} \circ \operatorname{grad}$, a differential operator on functions on a Riemannian manifold M. Then you can check that Δ commutes with $\operatorname{Isom}(M)$, and

$$\Delta = \frac{1}{\sqrt{g}} \sum_{i,j} \frac{\partial}{\partial x_i}\left(\sqrt{g}\, g^{ij} \frac{\partial}{\partial x_i}\right).$$

If S is a symmetric space, then you always have $\Delta \in \mathscr{D}(S)$.

Exercise 2.2.7. Show that for $\mathbf{R}^n$, with isometry group $G = \mathrm{Isom}(\mathbf{R}^n)$, we have $D(\mathbf{R}^n) = \mathrm{Poly}(\Delta)$.

Proposition 2.2.8. *The ring $\mathscr{D}(S)$ is finitely generated.*

Proof. We will use symmetrization. Let $f(x)$ be a function on S. For a basepoint x_0, define

$$f(x; x_0) = \int_{K_{x_0}} f(kx)\, dk. \tag{2.2.1}$$

Then evidently

$$f(\sigma x; x_0) = f(x; x_0) \text{ for } \sigma \in K_{x_0}.$$

We will warm up by reviewing a proof of the mean value theorem for harmonic functions. The same die will at the core of several important proofs.

We aim to show that for any harmonic function f, the value $f(x_0)$ agrees with the average of its values on a spherical shell around x_0. Consider the symmetrization (2.2.1) of f_0 – it is harmonic because Laplacian commutes with the symmetries. It is obviously radially symmetric. Since the Laplacian is a second-order differential equation, there should be 2 linearly independent solutions: c and $\log r$. But $\log r$ blows up as $r \to 0$, so it cannot be involved.

Since G is transitive, any $D \in \mathscr{D}(S)$ is determined by its action on the germ of functions at a fixed point x_0. If $L \in \mathscr{D}(S)$, then

$$[L \cdot f(x)]_{x=x_0} = [L \cdot f(x; x_0)]_{x=x_0}.$$

If $\mathscr{L}$ is any differential operator we can define an invariant $L \in \mathscr{D}(S)$ associated with it, by

$$Lf(x_0) = [\mathscr{L} f(x; x_0)]_{x=x_0}.$$

Why is this useful? We want to understand the relation between L and $\mathscr{L}$. For example, when do two $\mathscr{L}$'s give the same L? To discuss this we need to introduce the *symbol*.

Choose coordinates $x^1, \ldots, x^n$ near x_0 so that $g_{ij} = \delta_{ij}$. Then define differential operators

$$(\xi_1, \ldots, \xi_n) = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right).$$

Any differential operator $\mathscr{L}$ of degree m has a symbol, which is a polynomial of degree m in the variables $\xi_1, \ldots, \xi_n$, encoding the degree m homogeneous part of $\mathscr{L}$. Moreover, if $\mathscr{L} = L$ is invariant then $P(\xi_1, \ldots, \xi_n)$ is H-invariant, where $H \subset O(n)$ is the image of K_{x_0}.

Therefore, the symbol map from invariant differential operators to polynomials lies in a graded component of the invariants of a finitely generated polynomial ring under a compact group action. Therefore the image is finitely generated. □

Exercise 2.2.9. Complete the proof.

2.3. **Spherical functions.** Thanks to Proposition 2.2.8, we can choose a finite generating set $D_1, \ldots, D_\ell$ for $\mathscr{D}(S)$, which we fix.

Definition 2.3.1. We say that f is an *eigenfunction* of $\mathscr{D}(S)$ with *eigenparameter* $\lambda = (\lambda_1, \ldots, \lambda_j)$ if

$$D_j f = \lambda_j f \quad j = 1, \ldots, \ell.$$

Since the $\{D_j\}$ generate $\mathscr{D}(S)$, any such f is an eigenfunction for all operators in $\mathscr{D}(S)$.

Remark 2.3.2. In particular, f is an eigenfunction of the Laplacian, and is automatically smooth since Δ is elliptic. If S if analytic, then f is moreover analytic.

Definition 2.3.3. A *spherical* (or zonal) function is an eigenfunction which is radially symmetric about x_0, i.e.

$$f(x) = f(x; x_0) = f(\sigma x; x_0) \text{ for all } \sigma \in K_{x_0}.$$

We say that x_0 is the *pole* of f.

Theorem 2.3.4 (Multiplicity 1)**.** *The space of spherical functions about x_0 with eigenparameter λ is at most 1-dimensional, and if it is non-vanishing then there is a unique zonal function $w_\lambda(x_0)$ such that $w_\lambda(x_0; x_0) = 1$.*

Proof. We will show that the Taylor series of any spherical function about x_0 is uniquely determined. Here we use the analyticity of S and of any eigenfunction for $\mathscr{D}(S)$ (by elliptic regularity).

At x_0 choose coordinates $(x^1, \dots, x^n)$ such that $g_{ij} = \delta_{ij}$. Suppose f is a spherical function. We will compute

$$\left[\underbrace{\left(\frac{\partial}{\partial x^1}\right)^{r_1} \cdots \left(\frac{\partial}{\partial x^n}\right)^{r_n}}_{\mathscr{L}} f \right]_{x=x_0} \tag{2.3.1}$$

Using that f is spherical, (2.3.1) agrees with

$$\left[\mathscr{L} f(x; x_0)\right]_{x=x_0} = [L f(x; x_0)]_{x=x_0} \tag{2.3.2}$$

where L be the symmetrization of $\mathscr{L}$. Now, we know that we can express $L = P_L(D_1, \dots, D_\ell)$ for some polynomial P_L, which necessarily acts as multiplication by $P_L(\lambda_1, \dots, \lambda_\ell)$. Hence

$$[L f(x; x_0)]_{x=x_0} = f(x_0).$$

If $f(x_0) = 0$ then all the coefficients are 0. If $f(x_0) \neq 0$, we can normalize it to be 1, and g is determined purely by λ. $\square$

We always use $w_\lambda(x; x_0)$ to denote the unique normalized non-zero zonal function (when it exists). Now we discuss the mean value property which is a generalization of the mean value property for harmonic functions. (As in the proof of Proposition 2.2.8, the core is the fact that there's only one solution smooth at the origin.)

Let f be an eigenfunction on S of $\mathscr{D}(S)$ with eigenparameter λ. Then

$$\int_{K_{x_0}} f(kx)\,dk = f(x; x_0) = f(x_0) w_\lambda(x; x_0). \tag{2.3.3}$$

Next we have a key lemma, which computes the eigenvalue of f under point-pair invariants.

Lemma 2.3.5. *Let f be an eigenfunction of $\mathscr{D}(S)$ with eigenparameter λ. Then*

$$\int_S k(x, y) f(y)\,dy = \widehat{k}(\lambda) f(x)$$

where $\widehat{k}(\lambda)$ is the Selberg (– Harish-Chandra) transform[8]

$$\widehat{k}(\lambda) = \int_S k(x_0, x) w_\lambda(x; x_0)\,dx$$

(which is independent of x_0).

[8] Also called the spherical transform.

Proof. Since $\int_S k(x_0,y)f(y)\,dy$ is symmetric about x_0, we have

$$\int_S k(x_0,y)f(y)\,dy = \int_S k(x_0,y)f(y;x_0)\,dy.$$

Then using (2.3.3) we find that

$$\int_S k(x_0,y)f(y;x_0)\,dy = f(x_0)\int_S k(x_0,y)w_\lambda(y;x_0)\,dy.$$

$\square$

Example 2.3.6. Consider $S=\mathbf{H}^2$, so $\mathscr{D}(S)=\mathrm{poly}(\Delta)$ where

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

We consider $(\Delta+\lambda)f=0$. You can check that

$$\Delta y^s + s(1-s)y^s = 0.$$

Thus y^s is a solution, with $\lambda=s(1-s)$. To make a spherical function, we average the function $F_\lambda(z) := y^s$. Choose $z_0=i$, which has stabilizer $K_i=\mathrm{SO}_2(\mathbf{R})$. Then

$$w_\lambda(z;i)=\int_{K_i} y(kz)\,dk.$$

Let $G=\mathrm{SL}_n(\mathbf{R})$, $\mathscr{Y}$ as before. We have the Cartan decomposition $G=NAK$, where

$$N=\begin{pmatrix}1 & * & * \\ & 1 & * \\ & & 1\end{pmatrix},\quad A=\begin{pmatrix}* & 0 & 0\\ 0 & * & 0\\ 0 & 0 & *\end{pmatrix},\quad K=\mathrm{SO}(n).$$

So any $g\in G$ can be written as

$$g=nak,\quad n\in N, a\in A, k\in K.$$

Let $\mathfrak{a}=\mathrm{Lie}(A)$. Write

$$g=ne^{H(a)}k,\quad a\in\mathfrak{a}. \tag{2.3.4}$$

For $\lambda\in\mathfrak{a}^*_{\mathbf{C}}$, define[9]

$$F_\lambda(g)=e^{\langle\lambda-\rho,H(g)\rangle}.$$

Here ρ is half the sum of the positive roots, and $H(g)=H(a)$ in the decomposition (2.3.4).

Proposition 2.3.7. *For $S=\mathrm{SL}_n(\mathbf{R})/\mathrm{SO}_n(R)$, $F_\lambda(g)$ is an eigenfunction of $\mathscr{D}(S)$. Take $x_0=\mathrm{Id}$. Then the unique spherical function with pole x_0 and eigenparameter $\lambda\in\mathfrak{a}^*_F$ is*

$$w_\lambda(g,I)=\int_K e^{\langle\lambda-\rho,H(kg)\rangle}\,dk.$$

[9]Why do we complexify? In the end we'll be trying to do some L^2-theory. We'll need functions which are defined in some region of $\mathfrak{a}^*_{\mathbf{C}}$, and they'll have to be analytically continued to realm of interest. This is the whole point of Eisenstein series.

Remark 2.3.8. What does all this have to do with representation theory? We are interested in $L^2(\Gamma\backslash G)$. What does it mean to restrict to K-invariant functions? Look at irreducible unitary representations $\rho: G \to U(H)$. We say that ρ is *spherical* or *unramified* if there exists a non-zero $v \in H$ which is K-invariant; this v is necessarily unique up to scalar.

Consider the matrix coefficient

$$\omega(g) = \langle \rho(g)v, v\rangle.$$

This is right-invariant by K, since v is fixed by K, but also left-invariant by K by unitarity. So this function is bi-invariant under K. It is the spherical function.

2.4. **Kernels.** We have a map

$$A(S) \to \mathrm{Op}(\mathrm{Cont}(S))$$

sending $k(x, y)$ to the operator K_k with kernel function $k(x, y)$. This is a ring homomorphism:

$$k_1 * k_2 \mapsto K_{k_1} \circ K_{k_2}.$$

Let G be a Lie group. We are interested in the "decomposition" of $L^2(G)$ into irreducible representations of G. These representations are by definition *tempered*.

If G is compact, then $L^2(G)$ decomposes into a direct sum of finite-dimensional representations. In general, the irreducibles are infinite-dimensional, and are not summands.

Example 2.4.1. Let $G = \mathbf{R}$. Then the Laplacian is d^2/dx^2, and its eigenfunctions are $e(x\zeta) = e^{2\pi i\zeta}$. These grow out of control at $\pm\infty$ unless $\zeta \in \mathbf{R}$, so the tempered functions are those with $\xi \in \mathbf{R}$.

In general the growth of the volume in a hyperbolic space is fast – exponential. So you need functions that decay quickly enough at the boundary. The tempered functions are those which are "almost" in L^2.

Let S be a symmetric space, and consider the Hilbert space $L^2(S)$. For $k \in A(S)$, the adjoint of K_k is K_{k^*} where $k^*(x, y) = \overline{k(y, x)} \in A(S)$. Hence by Lemma 2.2.5, K_k is *normal*: it commutes with its adjoint. This is what will allow us to simultaneously diagonalize our operators.

Now we consider the situation for locallly symmetric spaces. Let $\Gamma \leq G = \mathrm{Isom}(S)$ be a discrete subgroup. Form the locally symmetric space $X_\Gamma := \Gamma\backslash S$. We are interested in the space

$$L^2(X_\Gamma) = \left\{ f : \Gamma\backslash S \to \mathbf{C} \mid \int_{X_\Gamma} |f(x)|^2 \, dx < \infty \right\}.$$

We define a homomorphism

$$A(S) \to \mathrm{Op}(L^2(X_\Gamma))$$

by averaging the kernel function over Γ in one argument:

$$k_\Gamma(x, y) = \sum_{\gamma \in \Gamma} k(\gamma x, y).$$

For $k \in A(S)$, $f \in C(\Gamma\backslash S)$, the operator $K_k \in \mathrm{Op}(L^2(X_\Gamma))$ is defined by

$$K_k f(x) = \int_{\gamma\backslash S} k_\gamma(x, y) f(y) \, dy.$$

An important trick is that this can be "unfolded":

$$K_k f(x) = \int_{\Gamma\backslash S} \sum_{\gamma\in\Gamma} k(x,\gamma y) f(y)\, dy$$
$$= \int_S k(x,y) f(y)\, dy.$$

This formula makes it clear that the map $k \mapsto K_k$ is a homomorphism. If we consider K_k as an operator on $L^2(X_\Gamma)$ then we can speak of adjoints, and $K_k^* = K_{k^*}$.

For simplicity we assume that X_Γ *is compact.* Then the algebra generated by K_k for $k \in A(S)$ consists of compact operators.

2.5. **Fredholm theory.**

Definition 2.5.1. Let H be a Hilbert space. An operator K on H is *compact* if K is a uniform limit of finite rank operators.

Suppose X is a locally symmetric space X_Γ. Let $k(x,y)$ be a kernel function on $X \times X$, and K the corresponding operator

$$(Kf)(x) = \int_X k(x,y) f(y)\, dy.$$

Definition 2.5.2. We say that an operator K associated to a kernel function k is *Hilbert-Schmidt* if

$$\int |k(x,y)|^2\, dx\, dy < \infty.$$

Theorem 2.5.3. *If K is Hilbert-Schmidt, then K is compact.*

We don't want to work with the Laplacian, which is an unbounded operator. The inverse of the Laplacian is an integral operator, which will tend to be compact.

Theorem 2.5.4 (Fredholm)**.** *Let H be a Hilbert space and $K: H \to H$ a compact operator. Then there are $\lambda_j \in \mathbf{C} - \{0\}$ such that $|\lambda_j| \to 0$, and such that the non-zero eigenspaces*

$$V_{\lambda_j} := \{v : Kv = \lambda_j v\}$$

are finite-dimensional. For $\lambda \notin \{\lambda_j\} \smile \{0\}$, the operator $R_\lambda(K) := (\lambda\,\mathrm{Id} - K)^{-1}$ is bounded.

In fact, Fredholm defines a version of the characteristic polynomial for such a K,

$$\Delta(\lambda) := \det(\mathrm{Id} - \lambda K).$$

Since there are infinitely many eigenvalues, this is an entire function (rather than a polynomial), whose reciprocal eigenvalues are the λ_j in Theorem 2.5.4).

Theorem 2.5.5 (Spectral theorem)**.** *If K is a compact operator on a Hilbert space H such that $KK^* = K^*K$ (i.e. K is normal), then there is an orthonormal basis $\phi_1, \phi_2, \ldots$ of H such that $K\phi_j = \lambda_j \phi_j$.*

Corollary 2.5.6. *Assuming X_Γ is compact, there is an orthonormal basis ϕ_j of joint eigenfunctions of $A(S)$ and $\mathscr{D}(S)$. Furthermore, each eigenspace of $\mathscr{D}(S)$ for an eigenparameter λ is finite-dimensional.*

Proof. We first show that the eigenspaces of the point-pair algebra $A(S)$ are finite-dimensional. As we've seen, for every $k \in A(S)$ the operators K_k are compact normal, and commute. Hence they can be simultaneously diagonalized. The only thing that has to be ruled out is that they have a large common kernel. This follows from the fact that we can find an approximate identity $k_\delta \to \mathrm{Id}$ in $A(S)$, whose eigenvalues tend to 1.

Next, we turn out attention to eigenspaces for the Laplacian. By Lemma 2.3.5 the eigenvalues of K_k, for any $k \in A(S)$, depend only on the eigenvalues of the Laplacian. Therefore the finite-dimensionality of the eigenspaces for $A(S)$ implies the finite-dimensionality of the eigenspaces for $\mathscr{D}(S)$.

□

In particular for an eigenfunction ϕ_λ of $\mathscr{D}(S)$, note that

$$K_k\phi_\lambda(x) = \widehat{k}(\lambda)\phi_\lambda(x)$$

where $\widehat{k}$ is the spherical transform of Lemma 2.3.5.

Theorem 2.5.7. *We have*

$$K_k(x,y) = \sum_{\gamma\in\Gamma} k(\gamma x, y) = \sum_j \widehat{k}(\lambda_i)\phi_j(x)\overline{\phi_j(y)} \tag{2.5.1}$$

where ϕ_i *has eigenparameter* $\lambda_j = (\lambda_j^{(1)},\dots,\lambda_j^{(\ell)})$. *In particular*

$$\mathrm{Tr}(K_k) = \sum_j \widehat{k}(\lambda_j). \tag{2.5.2}$$

Remark 2.5.8. The spectral expansion of the kernel is (2.5.1). The equation (2.5.2) is the *spectral side of the trace formula.*

Proof. Fix x, and expand $K(x,y)$ viewed as a function in y in $L^2(X_\Gamma)$:

$$K(x,y) = \sum_j \langle K(x,-),\phi_j\rangle\overline{\phi_j(y)}.$$

We then compute the coefficients by unfolding:

$$\begin{aligned}\langle K_k(x,\cdot),\phi_j\rangle &= \int_{\Gamma\backslash S} k_\Gamma(x,y)\overline{\phi}_j(y)\,dy\\ &= \int_S k(x,y)\phi_j(y)\,dx\\ &= \widehat{k}(\lambda_j)\phi_j(x).\end{aligned}$$

□

3. Eisenstein series

3.1. **Hyperbolic space.** Consider the symmetric space $S = \mathbf{H}$ (2-dimensional hyperbolic space, which is also denoted $\mathbf{H}^2$). Then $\mathscr{D}(\mathbf{H})$ is the ring of polynomials in Δ, and $A(\mathbf{H})$ is the ring of compactly supported functions in the distance function $d(z,\zeta)$. This is a symmetric space for $G = \mathrm{SL}_2(\mathbf{R})$, i.e we have $\mathbf{H} = \mathrm{SL}_2(\mathbf{R})/\mathrm{SO}_2(\mathbf{R})$, with $g \in \mathrm{SL}_2(\mathbf{R})$ acting by linear fractional transformation, and preserving the hyperbolic metric

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

In fact **H** is also a complex space, and G acts by biholomorphic maps. Let $\Gamma \leq \mathrm{SL}_2(\mathbf{R})$ be a discrete subgroup, and X_Γ the attached Riemann surface. We are interested in $\Gamma = \mathrm{SL}_2(\mathbf{Z})$. For this Γ, $\Gamma\backslash\mathbf{H}$ is a modular surface of finite area, which is *not* compact.

To see why, recall some reduction theory. The group Γ is generated by

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

These act by $z \mapsto -1/z$ and $z \mapsto z+1$, respectively, on **H**.

The Laplacian on **H** is

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

For $s \in \mathbf{C}$, we have

$$\Delta y^s + s(1-s)y^s = 0.$$

We want to make this into a function on $\Gamma\backslash\mathbf{H}$.

Note that

$$\operatorname{Im}(\gamma z) = \frac{y}{|cz+d|^2},$$

so in particular $\operatorname{Im}(Sz) = \frac{y}{|z|^2}$.

Using T we can bring any $z \in \mathbf{H}$ to the region $-1/2 \leq \operatorname{Re}(z) \leq 1/2$. We claim we can further move z to the domain $|z| \geq 1$. Indeed, using S (and T) we can keep finding orbits with increasing y-coordinate. Either we land in $\{x^2 + y^2 \geq 1\}$ or we have an infinite sequence of points tending to the boundary. By discreteness of $\mathrm{SL}_2(\mathbf{Z})$, it acts discontinuously on **H**, hence its orbits have no limit points.

The fundamental case is $\Gamma = \mathrm{SL}_2(\mathbf{Z})$. For any finite-index subgroup $\Gamma < \mathrm{SL}_2(\mathbf{Z})$, it is clear that X_Γ has finite area.

Definition 3.1.1. We say that Γ is a *congruence subgroup* of $\mathrm{SL}_2(\mathbf{Z})$ if $\Gamma(N) \subset \Gamma \subset \mathrm{SL}_2(\mathbf{Z})$ for some N, where

$$\Gamma(N) = \{\gamma \in \mathrm{SL}_2(\mathbf{Z}) \colon \gamma \equiv \mathrm{Id} \mod N\}.$$

Exercise 3.1.2. Show that $\mathrm{SL}_2(\mathbf{Z})$ has noncongruence subgroups of finite index.

3.2. **Some applications.** Why study the spectrum of Δ? We will give some fun examples.

Let $d(n) := \sum_{d|n} 1$. Consider the Dirichlet series

$$\sum_{n=1}^{\infty} \frac{d(n)}{n^s}.$$

It can be analyzed using the identity

$$\sum_{n=1}^{\infty} \frac{d(n)}{n^s} = \zeta(s)^2.$$

This can also be viewed as the L-function of an Eisenstein series.

What about

$$\sum \frac{d(n)^2}{n^s}?$$

Use the identity

$$\sum \frac{d(n)^2}{n^s} = \frac{\zeta(s)^4}{\zeta(s)}.$$

Now let

$$\sigma_{it}(n) = \sum_{d|n} d^{it},$$

imagining that t is real. Then

$$\sum_{n=1}^{\infty} \frac{|d_{it}(n)|^2}{n^s} = \frac{\zeta^2(s)\zeta(s+it)\zeta(s-it)}{\zeta(2s)}. \tag{3.2.1}$$

This leads to a quick proof of the prime number theorem, as follows. By the standard argument, the key point is to show that $\zeta(1+it) \neq 0$. By the symmetry of $\zeta(s)$ under complex conjugation, there would also be a zero at $\zeta(s-it)$. Then the function in (3.2.1) would be entire. But the left side is a Dirichlet series with positive coefficients, hence should have radius of convergence equal to the distance to the nearest pole.

How about the Dirichlet series

$$\sum_{n=1}^{\infty} \frac{d(n)d(n+h)}{n^s}?$$

It is a remarkable fact this has a meromorphic continuation to $\mathbf{C}$ with poles at $1/2 + it_j$, where $\lambda_j = 1/4 + t_j^2$ are the discrete eigenvalues of Δ on $L^2(X_{\mathrm{SL}_2(\mathbf{Z})})$.[10]

3.3. **Eisenstein series.** From now on let $X_\Gamma := \Gamma \backslash \mathbf{H}$. This has finite area but is not necessarily compact or arithmetic. For simplicity we assume there is *only one cusp*, which has stabilizer

$$\Gamma_\infty = \left\{ \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} : m \in \mathbf{Z} \right\}.$$

Note that y^s is a already a Γ_∞-invariant eigenfunction of Δ.

Definition 3.3.1. Define the *Eisenstein series*

$$E(z,s) := \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} y(\gamma z)^s.$$

Proposition 3.3.2. *The function $E(z,s)$ enjoys the following properties.*

(1) *It converges for* Re $s > 1$.
(2) $E(z,s) = E(\gamma z, s)$ *for* $\gamma \in \Gamma$.
(3) $\Delta E(z,s) + s(1-s)E(z,s) = 0$.
(4) $E(z,s) = y^s + O(1)$ *for s fixed and $y \to \infty$.*

Remark 3.3.3. Note that the Haar measure is $\frac{dx\,dy}{y^2}$. Are these functions $E(z,s)$ square-integrable? We need

$$\int_A^\infty |f(x,y)|^2 \frac{dx\,dy}{y^2} < \infty.$$

Hence the growth needs to be $o(y^{1/2})$. In other words, $E(z,s)$ for Re $s = 1/2$ just barely misses being L^2. But our series is only defined for Re $s > 1$, so this is why we need to prove meromorphic continuation.

Proof. We prove (1). Choose $\delta > 0$ very small so that the images of $B(z_0,\delta)$ under Γ are essentially disjoint. Then by the mean value property

$$\int_{B(z_0,\delta)} y^\sigma \frac{dx\,dy}{y^2} = \frac{y^\sigma}{c(\delta,\sigma)}.$$

[10]The poles are independent of h, but the residues depend on h.

We can rewrite this as

$$y^\sigma = c(\delta,\sigma)\int_{B(z_0,\delta)} y^\sigma \frac{dx\,dy}{y^2}.$$

Hence

$$\sum_{\gamma\in\Gamma_\infty\backslash\Gamma} y(\gamma z)^\sigma = y^\sigma + c(\delta,\sigma)\int_{\bigcup_{\gamma\in\Gamma_\infty\backslash\Gamma-\mathrm{Id}} B(\gamma z_0,\delta)} y^\sigma \frac{dx\,dy}{y^2}.$$

Replacing the union with a part of the fundamental domain, it is bounded by

$$\int_{x=0}^{1}\int_{y=0}^{h(x)} y^\sigma \frac{dx\,dy}{y^2},$$

which is finite.

The other parts follow easily.

□

Since $(\Delta + s(1-s))E(z,s)=0$, the residue at $s=1$ is a harmonic function, hence is constant. What is this constant?

Proposition 3.3.4. *We have*

$$\mathrm{Res}_{s=1} E(z,s) = \frac{1}{\mathit{Area}(\Gamma\backslash\mathbf{H})}.$$

Remark 3.3.5. If we can compute this residue in another way, we can use Proposition 3.3.4 to *compute* the volume of $\Gamma\backslash\mathbf{H}$. This is Langlands' idea to compute the Tamagawa number of algebraic groups. In general, for a split group G this relates the volume of $\Gamma\backslash G$ and the volume of $M\cap\Gamma\backslash M$ where M is a Levi. This allows to inductively compute the Tamagawa numbers of split groups.

Proof. Since y^s is holomorphic at $s=1$, we have

$$\mathrm{Res}_{s=1} E(z,s) = \mathrm{Res}_{s=1}[E(z,s)-y^s].$$

By the constancy of $\mathrm{Res}_{s=1} E(z,s)$ in z, mentioned above, we have for a fundamental domain $\mathscr{Y}$ of $\Gamma\backslash\mathbf{H}$,

$$\begin{aligned}
\mathrm{Res}_{s=1}[E(z,s)-y^s] &= \frac{1}{\mathrm{Vol}(\mathscr{Y})}\int_{\mathscr{Y}} \mathrm{Res}_{s=1}[E(z,s)-y^s]\frac{dx\,dy}{y^2}\\
&= \frac{1}{\mathrm{Vol}(\mathscr{Y})}\mathrm{Res}_{s=1}\int_{\mathscr{Y}}[E(z,s)-y^s]\frac{dx\,dy}{y^2}\\
&= \frac{1}{\mathrm{Vol}(\mathscr{Y})}\mathrm{Res}_{s=1}\int_{\mathscr{Y}}\sum_{\gamma\in\Gamma_\infty\backslash\Gamma-\mathrm{Id}} y(\gamma z)^s\frac{dx\,dy}{y^2}
\end{aligned}$$

Now unfold to a fundamental domain $\mathscr{F}$ for $\Gamma_\infty\backslash\mathbf{H}$:

$$\begin{aligned}\frac{1}{\mathrm{Vol}(\mathscr{Y})}\mathrm{Res}_{s=1}\int_{\mathscr{Y}}\sum_{\gamma\in\Gamma_\infty\backslash\Gamma-\mathrm{Id}} y(\gamma z)^s\frac{dx\,dy}{y^2} &= \frac{1}{\mathrm{Vol}(\mathscr{Y})}\mathrm{Res}_{s=1}\int_{\mathscr{F}} y^s\frac{dx\,dy}{y^2}\\ &= \frac{1}{\mathrm{Vol}(\mathscr{Y})}\mathrm{Res}_{s=1}\int_0^1\int_0^{h(x)} y^{s-2}\,dy\,dx\\ &= \frac{1}{\mathrm{Vol}(\mathscr{Y})}\mathrm{Res}_{s=1}\int_0^1\frac{h(x)^{s-1}}{s-1}\,dx\\ &= \frac{1}{\mathrm{Vol}(\mathscr{Y})}.\end{aligned}$$

□

3.4. **Fourier expansion.** We assume that we are in the domain of convergence Re $s\geq 1$. Since

$$E(z+\ell,s)=E(z,s)$$

we have a Fourier expansion, writing $z=x+iy$,

$$E(z,s)=\sum_m a_m(s,y)e(mx).$$

We will study the Fourier coefficients.

Separating variables, using $\Delta E(z,s)+s(1-s)E(z,s)=0$, we find that $a_m(s,y)$ satisfies an ordinary differential equation

$$y^2a_m''(y)+s(1-s)a_m(y)-y^2 4\pi^2m^2a_m(y)=0.$$

If we make the change of variables $b_m(y)=y^{1/2}a_m(y)$, then $b_m(y)$ solves the differential equation

$$b_m''(y)+y\,b_m'(y)+\left[\frac{-1/4-?(s)}{y^2}-4\pi^2m^2\right]b_m(y)=0.$$

There is a 2-dimensional space of solutions. We want to think about the behavior as $y\to\infty$ (much as we did when reasoning about the mean value property). If $m\neq 0$, there is a 1-dimensional solution space which decays (exponentially) at ∞ and a 1-dimensional space that grows exponentially. On the other hand, by Proposition 3.3.2 we know that $E(z,s)=y^s+O(1)$ as $y\to\infty$. Since

$$\int|E(z,s)|^2\,dx=\sum|a_m(y)|^2$$

we must have $|a_m(y)|=O(y^{2\sigma})$. Hence the mth coefficient of $E(z,s)$ must pick the exponentially decaying solution.

You can check that for $m\neq 0$, this solution is

$$a_m(y)=a_m y^{1/2}K_{s-1/2}(2\pi|m|y).$$

For $m=0$, the solutions are y^s, y^{1-s} except at $s=1/2$, you get y^s and $y^s\log y$. Therefore

$$\int_0^1 E(z,s)\,dx=\alpha_s y^s+\beta_s y^{1-s}.$$

From the known behavior of $E(z,s)$ as $y\to\infty$, we know that $\alpha=1$. Selberg called $\phi(s):=\beta_s$. This is holomorphic for Re $s\gg 1$.

We have concluded the expansion

$$E(z,s) = y^s + \phi(s)y^{1-s} + \sum_{m\neq 0} a_m y^{1/2} K_{s-1/2}(2\pi|m|y)e(mx). \tag{3.4.1}$$

In particular,

$$E(z,s) = y^s + \phi(s)y^{1-s} + O_s(e^{-cy}) \text{ for some } c>0.$$

Since y is bounded away from 0, the concern for square-integrability is at $y\to\infty$. Since we have to balance y^s and y^{1-s}, we want to be on the line $\mathrm{Re}\ s = 1/2$. It's also an issue that $\phi(s)$ might have a pole – this has to do with the famous "residues of Eisenstein series". Hence this suggests that to get to the $L^2(X_\Gamma)$ continuous spectrum, we need to meromorphically continue $E(z,s)$ to the line $\mathrm{Re}\ s = 1/2$.[11]

Exercise 3.4.1. Show that the spectrum of Δ is partly continuous. Hint: start with the function $y^{1/2+it}$, for $t\in\mathbf{R}$ such that $s(1-s) = 1/4+t^2$. This isn't L^2, so we use a cutoff function $\psi_{A,t}(z) = \eta(y)y^{1/2+it}$ where $\eta(y)$ is a smooth function supported in $(A,2A)$ so that

$$\|\Delta\psi_{A,t} - (1/4+t^2)\psi_{A,t}\| \to 0 \text{ as } A\to\infty$$

while $\|\psi_{A,t}\|_2 = 1$.

Corollary 3.4.2. *The spectrum of* $\Delta|_{L^2(X_\Gamma)}$ *contains* $[1/4,\infty)$.

Exercise 3.4.3. If $\Gamma = \mathrm{SL}_2(\mathbf{Z})$, show that

$$E(z,s) = y^s + \frac{\Lambda(2s-1)}{\Lambda(2s)}y^{1-s} + \sum_{n=1}^{\infty}\frac{2\sigma_{s-1/2}(m)}{\Lambda(2s)}y^{1/2}K_{s-1/2}(2\pi m y)\cos(2\pi m x)$$

where

$$\Lambda(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$$

is the completed ζ function, hence $\phi(s) = \frac{\Lambda(2s-1)}{\Lambda(2s)}$, and

$$\sigma_w(n) = \sum_{d|n} d^w.$$

Hint: use a coset decomposition

$$\Gamma_\infty\backslash\Gamma \cong \{\begin{pmatrix} * & * \\ c & d\end{pmatrix} : (c,d)=1\}.$$

Here is an interesting application, due to Deuring.

Theorem 3.4.4 (Deuring)**.** *Suppose that the Riemann Hypothesis is false. Then* $h(Q(\sqrt{-D})) = 1$ *for only finitely many* D.

Proof. We use the fact that there are $h(D)$ inequivalent binary quadratic forms of discriminant $D<0$. Let $E^*(z,s) = \Lambda(2s)E(z,s)$. Then we have the identity

$$\zeta_{\mathbf{Q}(\sqrt{-D})}(s) = \sum_{z_j\in \mathrm{CM}(-D)} E^*(z_j,s).$$

On the other hand, we have

$$\zeta(s)L(s,\chi_{-D}) = \zeta_{\mathbf{Q}(\sqrt{-D})}(s).$$

[11]In Langlands' normalization, this is $\mathrm{Re}\ s = 0$.

Suppose $h(-D)=1$. The principal class is always attached to the CM point $z_1=-\frac{1}{2}+\frac{i\sqrt{D}}{2}$. Consider the identity

$$E(z_1,s)=\zeta(s)L(s,\chi_{-D}).$$

Let ρ be a zero of $\zeta(s)$ with $\operatorname{Re} s>1/2$. Then $E^*(z_1,\rho)=0$, but using the Fourier expansion (3.4.1) we also have

$$E^*(z_1,\rho)=\Lambda(2\rho)(\sqrt{D})^{\rho}+\Lambda(2\rho-1)(\sqrt{D})^{1-\rho}+O(e^{-c\sqrt{D}}).$$

Since the error term goes to 0, this is only possible if $\rho=1/2$ since the first term dominates. $\square$

Remark 3.4.5. This idea is used to solve the "class number problem", which is to find for fixed m which D have $h(D)=m$. An effective solution was provided by Goldfeld, using the Gross-Zagier formula – it yields

$$h(D)\geq\frac{\log|D|}{10^{120}}.$$

Obviously, this isn't very practical. One would like to have a method to bring the bound down from 10^{120} to something like 10^{10}. Such a method was given by S. Arno, which allows to check intermediate values. The idea is to examine the equation

$$E(z,s)=y^s+\phi(s)y^{1-s}+o(e^{-cy})$$

and to use the zeros of ζ to deduce small relations among the imaginary parts.

4. Analytic continuation of Eisenstein series

4.1. **Analysis of the kernel function.** We now consider a general lattice $\Gamma\subset \mathrm{SL}_2(\mathbf{Z})$. The most important thing is to understand the behavior of the point-pair kernel as you go into the cusp.

Let $k\in A(\mathbf{H})$, and recall that we defined k to have compact support[12]. Let

$$k_\Gamma(z,s)=\sum_{\gamma\in\Gamma}k(\gamma z,s).$$

We study the attached kernel K_Γ on $\Gamma\backslash\mathbf{H}$. On $\mathbf{H}$, all point-pair invariants $k(z,\zeta)$ are functions in the hyperbolic distance function,

$$d(z,\zeta)=\frac{|z-\zeta|^2}{y\eta}\qquad z=x+iy,\zeta=\xi+i\eta.$$

Let $k(z,\zeta)=\Phi(d(z,\zeta))$, so Φ has compact support. A key identity is the unfolding property

$$\int_{\Gamma\backslash\mathbf{H}}k_\Gamma(z,\eta)f(\zeta)d\zeta=\int_{\mathbf{H}}k(z,\zeta)f(\zeta)d\zeta. \tag{4.1.1}$$

If $f(\zeta)$ is an eigenfunction of Δ on $\Gamma\backslash\mathbf{H}$, say with eigenvalue $\lambda=s(1-s)$, then by Lemma 2.3.5,

$$\int_{\mathbf{H}}k(z,\zeta)f(\zeta)d\zeta=\widehat{k}(s(1-s))f(z)$$

where

$$\widehat{h}_k(s(1-s))_i=\int_{\mathbf{H}}k(i,\zeta)\eta^s\,\frac{d\xi\,d\eta}{\eta^2}.$$

[12] In practice this is occasionally relaxed to allow k to decay rapidly (basically because compact support is not preserved by Fourier transform).

We can write

$$\frac{|z-\zeta|^2}{y\eta} = \frac{(x-\xi)^2}{y\eta} + \frac{y}{\eta} + \frac{\eta}{y} - 2.$$

We want to study $k_\Gamma(z,\zeta)$ as a function on $\Gamma\backslash\mathbf{H}\times\Gamma\backslash\mathbf{H}$ when z or $\zeta\to\infty$ in the cusp. In particular we want to understand why it is not Hilbert-Schmidt.

If z is fixed and $\eta\to\infty$, then $k_\Gamma(z,\zeta)\equiv 0$, and vice versa. The problem is when z,ζ are both going to ∞ at "the same pace". So imagine that y and η both get large; the idea is that the only Γ-translates of z which are close to ζ are the translates under T, since the metric is $\frac{dz}{y^2}$:

$$\sum_{\gamma\in\Gamma} k(\gamma z,\zeta) = \sum_{m\in\mathbf{Z}} k(z+m,\zeta).$$

Now we use Poisson summation. (In general the sum would be over a unipotent radical, and you can still apply Poisson summation.)

Remark 4.1.1. When do you use Poisson summation? Consider a sum

$$\sum_n f(nt)$$

with f decaying rapidly (imagine that everything has compact support). Poisson summation says

$$\sum_n f(nt) = \frac{1}{t}\sum_m \widehat{f}(m/t).$$

If t is large, then the spacing is large. In that case you shouldn't apply Poisson summation, since it will transform the large sum to a short one.

In summary: if t is large, then leave it alone. If t is small, dualize. This basically amounts to replacing a sum by an integral.

Write

$$\psi_{y,\eta,x,\xi}(t) = \Phi\left(\frac{(x-\xi+t)^2}{y\eta} + \frac{y}{\eta} + \frac{\eta}{y} - 2\right).$$

Since y,η are getting large, this is becoming a short sum, so we should dualize. The Fourier transform is

$$\widehat{\psi}(\tau) = e^{2\pi i \frac{x-\xi}{\sqrt{y\eta}}\tau}(y\eta)^{1/2}\int_{\mathbf{R}} \Phi(u^2 + \frac{y}{\eta} + \frac{\eta}{y} - 2)e^{2\pi i\sqrt{y\eta}\tau u}\,du.$$

The upshot is that we have *uniformly in* $\eta, y \geq 1/2$, a contribution from $\tau = 0$ in the Poisson sum plus negligible stuff.

$$\begin{aligned}\sum_{m\in\mathbf{Z}} k(z,\zeta+m) &= \sum_{m\in\mathbf{Z}} \psi_{y,\eta,x,\xi}(m)\\ &= (y\eta)^{1/2}\int_{\mathbf{R}} \Phi(u^2+\frac{y}{\eta}+\frac{\eta}{y}-2)e^{2\pi i\sqrt{y\eta}\tau u}\,du + O_N((y\eta)^{-N})\\ &= \int_{\mathbf{R}} k(z,\zeta+t)\,dt + O((y\eta)^{-N}).\end{aligned}$$

The key estimate is:

$$\sum_{m\in\mathbf{Z}} k(z,\zeta+m) = \int_{\mathbf{R}} k(z,\zeta+t)\,dt + O((y\eta)^{-N}). \tag{4.1.2}$$

To summarize, we have found that

$$k_\Gamma(z,\zeta) = \begin{cases} 0 & \frac{y}{\eta} + \frac{\eta}{y} \geq A_2 = A_2(k) \\ \sim (y\eta)^{1/2} + O((y\eta)^{-N}) & \text{otherwise.} \end{cases} \tag{4.1.3}$$

In particular, we see explicitly that $k_\Gamma(z,\zeta)$ is not Hilbert-Schmidt, as expected (since the spectrum is not discrete).

What's coming next? We're going to try to use Fredholm theory to analytically continue. For $\operatorname{Re} s > 1$, $E(z,s)$ makes sense and so does

$$K_\Gamma(E(z,s)) = \int_{\mathbf{H}} k(z,\zeta)E(\zeta,s)\,d\zeta = \widehat{k}(s(1-s))E(z,s).$$

We'll eventually express the Eisenstein series as a Fredholm determinant and deduce analytic continuation in that way.

Lemma 4.1.2 (Uniqueness principle)**.** *If f is an eigenfunction of Δ on $\Gamma\backslash\mathbf{H}$ of moderate growth, with eigenvalue $\lambda = s(1-s)$ for $\operatorname{Re} s > 1$, then $f(z) = \alpha E(z,s)$ for some constant α.*

Proof. We have seen that moderate growth for f implies that $f(z) = \alpha y^s + \beta y^{1-s} + O(1)$ as $y \to \infty$. (Here is where moderate growth is used. The argument is the same as for the Eisenstein series, §3.4.) So define

$$F(z) = f(z) - \alpha E(z,s) = \beta' y^{1-s} + O(1) \text{ as } y \to \infty.$$

So $F \in L^2(X_\Gamma)$, and is an eigenfunction of Δ, so its eigenvalue is real (self-adjointness) and positive (integration by parts). But $s(1-s)$ is not real and positive if $\operatorname{Re} s > 1$, we must have $F = 0$. □

Note the importance of using L^2-theory, which was introduced by Selberg. Earlier, people had thought of modular forms as living in finite-dimensional spaces.

4.2. **Cutoff functions.** Let $\alpha_A(y)$ be a smooth function which is 1 if $y \geq A$ and 0 if $y \leq A-1$, thinking A to be large. Now define

$$\alpha_A(z) = \begin{cases} \alpha_A(y) & y \geq A-1, \\ 0 & \text{otherwise} \end{cases}$$

viewed as an automorphic function. Define

$$\widetilde{E}(z,s) = E(z,s) - \alpha_A(z)y^s.$$

Since y^s is also an eigenfunction of Δ with eigenvalue $s(1-s)$ this is an eigenfunction with eigenvalues $s(1-s)$ off a compact set. By the asymptotic of Proposition 3.3.2, we will have $\widetilde{E}(z,s) \in L^2(X_\Gamma)$, and

$$K(\widetilde{E}(z,s)) = \widehat{k}(s(1-s))\widetilde{E}(z,s) + G(z,s)$$

where

$$G(z,s) = K(\alpha(z)y^s) - \widehat{k}(s(1-s))\alpha(z)y^s.$$

has compact support and extends to an entire function. We have (still assuming $\operatorname{Re} s > 1$)

$$(K - \widehat{k}(s(1-s)))\widetilde{E}(z,s) = G(z,s).$$

If we can invert $(K - \widehat{k}(s(1-s)))$ then we win. But we know this is problematic because $(K - \widehat{k}(s(1-s)))$ is not compact. We will modify it to be compact. This forces us to break self-adjointness.

One proof is to use the resolvent to prove analytic continuation to the line $\operatorname{Re} s = 1/2$, using a compactness argument to deduce finitely many poles in $[1/2, 1]$. This was Selberg's first proof, but it is very complicated.

We will give a different proof, by cutting off K. Define a smooth function $\widetilde{k}(z,\zeta)$ on $\Gamma\backslash\mathbf{H}\times\Gamma\backslash\mathbf{H}$ by:

$$\widetilde{k}(z,\zeta) = k(z,\zeta) - \underbrace{\alpha(z)\int_{\mathbf{R}} k(z,\zeta+t)\,dt}_{=:k_0(z,\zeta)}. \tag{4.2.1}$$

The idea is that if z is large, the two things are trying to behave the same. From the analysis of $k(z,\zeta)$ in the cusp (4.1.3), $\widetilde{k}(z,\zeta)$ is rapidly decreasing (in each variable) as $z,\zeta\to\infty$.[13] So $\widetilde{k}$ is Hilbert-Schmidt, hence gives a compact operator on $L^2(X_\Gamma)$. We can then apply Fredholm theory.

Remark 4.2.1. There is a natural subspace of $L^2_{\text{cusp}}(X_\Gamma) \subset L^2(X_\Gamma)$, called the *cuspidal* subspace. It is defined to be the subspace of f such that

$$\int_0^1 f(x+iy)\,dx = 0 \text{ for a.e. } y.$$

Exercise 4.2.2. Show that $L^2_{\text{cusp}}(X_\Gamma)$ is invariant under $A(H)$.

Note that $\widetilde{K}|_{L^2_{\text{cusp}}(X_\Gamma)} = K|_{L^2_{\text{cusp}}(X_\Gamma)}$ because the correction term depends on the constant term. Hence the compactness of $\widetilde{K}$ immediately gives:

Corollary 4.2.3. *The restriction $K|_{L^2_{\text{cusp}}(X_\Gamma)}$ is compact.*

We can then apply the theory of compact operators to deduce that the spectrum of $L^2_{\text{cusp}}(X_\Gamma)$ is discrete.

4.3. **Meromorphic continuation.** We now apply our operator $\widetilde{K}$, with kernel $k(z,\zeta) - k_0(z,\zeta)$ to $E(z,s)$. Since $k_0(z+x,\zeta) = k_0(z,\zeta)$, the term k_0 only picks up the 0th Fourier coefficient of $E(z,s)$. Hence we have

$$\widetilde{K}(E(z,s)) = \widehat{k}(s(1-s))E(z,s) - K_0(y^s + \phi(s)y^{1-s})$$

where

$$\begin{aligned} K_0(y^s) &= \alpha_A(y)\int k(z,\zeta)\eta^s \frac{dz\,d\eta}{\eta^2} \\ &= \alpha_A(y)\widehat{k}(s(1-s))y^s. \end{aligned}$$

Hence we have

$$(\widetilde{K} - \widehat{k}(s(1-s)))E(z,s) = -\alpha_A(y)\widehat{k}(s(1-s))(y^s + \phi(s)y^{1-s}). \tag{4.3.1}$$

This looks like a resolvent. We could try to use Fredholm theory, but there are two problems.

(1) The right side is not in L^2.
(2) The right side involves $\phi(s)$, so we would need to have control over $\phi(s)$. This shows that meromorphic continuation of $E(z,s)$ amounts to meromorphic continuation of $\phi(s)$. (The slogan is that "meromorphic continuation of Eisenstein series is equivalent to the meromorphic continuation of the constant term.")

[13] Note that although the first term $k(z,\zeta)$ is self-adjoint, the second term $k_0(z,\zeta)$ is certainly not.

The first issue (1) is easily solved by cutting off the Eisenstein series, since $\widetilde{k}(z,\zeta)$ is rapidly decreasing. The second issue (2) is more difficult to deal with.

We write down a key auxiliary equation. Let $E^*(z,s)$ solve the integral equation

$$[\widetilde{K}-\widehat{k}(s(1-s))]E^*(z,s)=-\alpha_A(y)\widehat{k}(s(1-s))y^s. \tag{4.3.2}$$

The RHS of (4.3.2) is not in L^2, but it is analytic in s, on *all* of $\mathbf{H}$. We will then use Fredholm theory to solve for $E^*(z,s)$. Set $E^{**}(z,s)=E^*(z,s)-\alpha(y)y^s$, which is defined for Re $s\gg 0$. Then $E^{**}(z,s)$ satisfies

$$(\widetilde{K}-\widehat{k}(s(1-s)))E^{**}(z,s)=\widetilde{K}(\alpha_A(y)y^s). \tag{4.3.3}$$

Now the RHS of (4.3.3) is in $L^2(\Gamma\backslash\mathbf{H})$, so we can apply Fredholm theory. We deduce that for s outside the spectrum of $\widetilde{K}$, we can define $E^{**}(z,s)$ in a big ball in the s parameter, as large as we please. Hence $E^{**}(z,s)$ extends to all of $\mathbf{C}$ as a meromorphic function for any fixed z, with poles *independent* of z. Hence $E^*(z,s)$ has a meromorphic continuation to $\mathbf{C}$.

Now we have to put $\phi(s)$ back. By the symmetry of the differential equation (4.3.2) with respect to $s\mapsto 1-s$, we have $E^*(z,s)=E^*(z,1-s)$ for all Re $s>1$. We then have

$$(\widetilde{K}-k(s(1-s)))(E^*(z,s)+\phi(s)E^*(z,1-s))=-\alpha_A(y)\widehat{k}(s(1-s))(y^s+\phi(s)y^{1-s}).$$

On the other hand, recall that

$$(\widetilde{K}-k(s(1-s)))(E(z,s))=\alpha_A(y)\widehat{k}(s(1-s))=-\alpha(y)\widehat{k}(s(1-s))(y^s+\phi(s)y^{1-s}).$$

The point is that this is the *same differential equation.* Now we apply a uniqueness principle.

Lemma 4.3.1. *For* Re $s>1$, $E(z,s)=E^*(z,s)+\phi(s)E^*(z,1-s)$.

Proof. For Re $s>1$, we know that $E^{**}(z,s)\in L^2(\Gamma\backslash H)$, hence $f(z):=E(z,s)-E^*(z,s)$ and satisfies

$$(\widetilde{K}-\widehat{k}(s(1-s)))F(z,s)=0.$$

We can choose s so that $\widehat{k}(s(1-s))$ is not in the spectrum of $\widetilde{K}$, and then we find that $F(z,s)\equiv 0$. □

Therefore, the analytic continuation of $E(z,s)$ is reduced to the analytic continuation of $\phi(s)$.

Lemma 4.3.2. *For* Re $s>1$*, and* $s\in\mathbf{R}$ *not a pole of* $E^*(z,s)$ *or* $E^*(z,1-s)$*, the function* $E^*(z,s)+\lambda E^{\lambda}(z,1-s)$ *is an eigenfunction for* Δ *with eigenvalue* $s(1-s)$ *if and only if* $\lambda=\phi(s)$.

Proof. If $\lambda\neq\phi(s)$, consider

$$f(z,s)=(E^*(z,s)+\phi(s)E^*(z,1-s))-(E^*(z,s)+\lambda E^*(z,1-s)).$$

Then $f(z,s)$ is an eigenfunction with eigenvalues $s(1-s)$ for Δ. For Re $s>1$, it is in L^2 and off the spectrum, hence $F=0$ and $\lambda=\phi(s)$. This gives an expression for λ with a visible analytic continuation. □

Consider the system of linear differential equations for $w\in\mathscr{Y}$ (the fundamental domain):

$$\Delta(E^*(w,s)+\lambda E^*(w,s))+s(1-s)(E^*(w,s)+\lambda E^*(w,1-s))=0.$$

Assuming Re $s>1$, by Lemma 4.3.2 this system is rank 1 and solvable in λ, i.e. there exists $w_0\in\mathscr{Y}$ (fundamental domain) where *system* is equivalent to the *single equation*

$$\Delta(E^*(w_0,s)+\lambda E^*(w_0,s))+s(1-s)(E^*(w_0,s)+\lambda E^*(w_0,1-s))=0.$$

Rewrite this as

$$\Delta(\lambda(s)E^*(w_0,s))=-s(1-s)(E^*(w_0,s)+\lambda E^*(w_0,1-s))-\Delta(E^*(w_0,s)$$

This presents $\lambda(s) = \phi(s)$ as the solution to a differential equation which has meromorphic continuation to all of $\mathbf{C}$. $\square$

4.4. **The functional equation.** Although we've now proved the meromorphic continuation, in practice we need to be able to control the function. We have control for Re $s > 1$, by the explicit equation, hence also for Re $s < 0$. To use a Phragmen-Lindelöf principle to control the critical strip, we need to prove that the order is finite. In fact, the order of *any* automorphic L-function should conjecturally be 1.

Theorem 4.4.1 (Functional equation)**.** *We have the following.*

(1) $E(z,1-s) = \phi(1-s)E(z,s)$.
(2) $\phi(s)\phi(1-s) = 1$.
(3) $|\phi(1/2+it)| = 1$, *and* ϕ *is analytic on* Re $s = 1/2$.

This gives us

$$E(z,1/2+it) = y^{1/2+it} + \phi(1/2+it)y^{1/2-it} + \dots.$$

Proof. Define

$$f(z,s) = \phi(1-s)E(z,s) - E(z,1-s).$$

The constant term is

$$\phi(1-s)y^s + \phi(1-s)\phi(s)y^{1-s} - y^{1-s} - \phi(1-s)y^s = \phi(1-s)\phi(s)y^{1-s} - y^{1-s}.$$

Hence for Re $s > 1$ we have $f(z,s) \in L^2(\Gamma\backslash\mathbf{H})$. We can choose s so that it is an eigenfunction off the spectrum of $\widetilde{K}$, hence $F = 0$. The equation $\phi(1-s)\phi(s) = 1$ follows by comparing the y^{1-s} terms. $\square$

5. The continuous spectrum

5.1. **Maass-Selberg relation.** For A large and fixed, make a cutoff so that

$$\widetilde{E}_A(z,s) = \begin{cases} E(z,s) - y^s - \phi(s)y^{1-s} & y \geq A, \\ E(z,s) & y < A. \end{cases}$$

We emphasize that we want this sharp cutoff – it is not continuous. We now want to compute the inner products of these functions for different values of s.

Lemma 5.1.1. *For* Re $s > 1$, *we have*

$$\int_{\mathscr{Y}} \widetilde{E}_A(z,s_1)\overline{\widetilde{E}_A(z,s_2)}\,dA(z) = \frac{A^{s_1+\overline{s_2}-1} - A^{1-s_1-\overline{s_2}}\phi(s_1)\overline{\phi(s_2)}}{s_1+\overline{s_2}-1} + \frac{\overline{\phi}(s_2)A^{s_1-\overline{s_2}} - \phi(s_1)A^{\overline{s_2}-s_1}}{s_1-\overline{s_2}}.$$

Proof. The proof works by cutting the region $\mathscr{Y}$ into pieces where the functions are smooth, and applying Stoke's theorem. We'll skip this calculation and focus on explaining qualitatively why the right hand side only depends on the constant term of the Eisenstein series. The point is that since we've only monkeyed around with the constant term, the answer can only depend on the constant term. $\square$

We now draw up two specific cases of Lemma 5.1.1. Appling Lemma 5.1.1 with $s_1 = s_2 = s = \sigma + ir$, we find that

$$\int_{\mathscr{Y}} |\widetilde{E}_A(z,s)|^2\,dA(z) = \frac{A^{2\sigma-1} - |\phi(\sigma-ir)|^2A^{1-2\sigma}}{2\sigma-1} + \frac{\overline{\phi(\sigma+ir)}A^{2ir} - \phi(\sigma+ir)A^{-2ir}}{2ir}. \quad (5.1.1)$$

Now fixing $r \neq 0$ and letting $\sigma \to 1/2$, we find

$$\int_{\mathscr{Y}} |\widetilde{E}_A(z,s)|^2 \, dA(z) = 2\log A - \frac{\phi'}{\phi}(1/2+ir) + \frac{\overline{\phi(1/2+ir)}A^{2ir} - \phi(1/2+ir)A^{-2ir}}{2ir}. \tag{5.1.2}$$

On the other hand, fixing $\sigma \neq 1/2$ and taking $r \to 0$ we find

$$\int_{\mathscr{Y}} |\widetilde{E}_A(z,s)|^2 \, dA(z) = \frac{A^{2\sigma-1} - \phi(\sigma)^2 A^{1-2\sigma}}{2\sigma-1} + \phi(\sigma)\log A - \phi'(\sigma). \tag{5.1.3}$$

Corollary 5.1.2. *The function $E(z,s)$ is holomorphic in* Re $s \geq 1/2$ *with only* simple *poles in* $[1/2,1]$.

Proof. Suppose $E(z,s)$ has a pole at $s=\rho$, with Re $\rho > 1/2$. We first want to show that the zeroes need to lie on the real line. Note that the poles of $E(z,s)$ are a subset of the poles of $\phi(s)$. So we have

$$E(z,s) = y^s + \phi(s)y^{1-s} + \dots$$

with $\phi(s)$ controlling poles of $E(z,s)$. If $\rho(1-\rho)$ is not real, then the leading pole term in $E(z,s)$, say $u(z)$, is an eigenfunction of Δ with eigenvalue $\rho(1-\rho)$, and is in L^2 because y^s is square-integrable for Re $s > 1/2$. But this contradicts the fact that Δ is self-adjoint. Thus the poles are in $(1/2,1]$.

Suppose ϕ has a pole of order > 1 at σ_0. Then the right hand side of (5.1.3) would be negative near σ_0, contradicting the obvious positivity of the left hand side. □

The residues of $E(z,s)$ at the poles $\rho_1,\dots,\rho_n$ are $u_i(z) \in L^2(\Gamma\backslash\mathbf{H})$, with eigenvalue $\rho_i(1-\rho_i)$. These are part of the discrete spectrum of $\Gamma\backslash\mathbf{H}$.

5.2. **Application.** Let $\Gamma = \mathrm{SL}_2(\mathbf{Z})$. The Eisenstein series is

$$E(z,s) = y^s + \phi(s)y^{1-s} + \frac{1}{\Lambda(2s)} \sum_{m\neq 0} \sigma_{s-1/2}(|m|) y^{1/2} K_{s-1/2}(2\pi|m|y) e(mx),$$

where $\phi(s) = \frac{\Lambda(2s-1)}{\Lambda(2s)}$, $\Lambda(s) = \pi^{-s/2}\Gamma(s)\zeta(s)$.

Let's use what we have developed to prove the prime number theorem. Suppose for $t_0 \in \mathbf{R}$, we have $\zeta(1+it_0) = 0$. Then the term $\frac{1}{\Lambda(1s)}$ would have a pole on Re $s = 1/2$, so by inspection of any non-zero Fourier coefficient in this Fourier expansion we find that $E(z,s)$ has a pole. But $\phi(s)$ has no pole on Re $s = 1/2$, contradiction.[14] □

5.3. **Continuous spectrum.** Using the theory of Eisenstein series, we determine the L^2-continuous spectrum. Let $X = \Gamma\backslash\mathbf{H}$, and let $\mathscr{Y}_\Gamma \subset \mathbf{H}$ be a fundamental domain for Γ. Let $f,g \in C_c^\infty(0,\infty)$ and consider

$$F(z) := \int_0^\infty f(r)E(z,1/2+ir)\,dr,$$

$$G(z) := \int_0^\infty g(r)E(z,1/2+ir)dr.$$

[14]In general, all proofs of non-vanishing come from this argument: examine a non-constant Whittaker coefficient for the Eisenstein series.

This is almost in L^2 since pointwise it is. Then we easily compute that

$$\int_0^\infty f(r)\overline{g(r)}\,dr = \frac{1}{2\pi}\int_{\mathscr{Y}_\Gamma} F(z)G(z)\,dA(z).$$

Hence the association $f \mapsto F$ induces an isometry from $L^2(0,\infty)$ to $L^2(\Gamma\backslash\mathbf{H})$. In this sense the Eisenstein series $E(z,1/2+ir)$ furnishes the L^2 continuous spectrum of $\Gamma\backslash\mathbf{H}$, and the eigenvalue of F is

$$\Delta F(z) = -\int_0^\infty f(r)(1/4+r^2)E(z,r)\,dr.$$

Lemma 5.3.1. *We have*

$$\langle F,G\rangle_{\mathscr{Y}_\Gamma} = \int_{\Gamma\backslash\mathbf{H}} F(z)\overline{G(z)}\,dA(z) = 2\pi\int_0^\infty f(r)\overline{g(r)}\,dr$$

Proof. Let A be a large parameter (for cutoff). We'll apply the Maass-Selberg relation, soset

$$E_A(z,1/2+it) = E(z,1/2+it) - \delta_A(y)(y^{1/2+it} + \phi(1/2+it)y^{1/2+it})$$

where $\delta_A(y)$ is a sharp cutoff

$$\delta(A)(y) = \begin{cases} 1 & y \ge A \\ 0 & y < A. \end{cases}$$

We then write

$$F(z) = \underbrace{\delta_A(y)\int_0^\infty (y^{1/2+ir} + \phi(1/2+ir)y^{1/2-ir})f(r)dr}_{F_1(z)} + \underbrace{\int_0^\infty E_A(z,1/2+ir)f(r)dr}_{F_2(z)}.$$

Using integration by parts, we find that $F_1(z) = y^{1/2}O(1/(\log y)^N)$ for any N. This estimate shows that $F_1(z) \in L^2(\mathscr{Y}_\Gamma)$, hence also $F_2(z) \in L^2(\mathscr{Y}_\Gamma)$. Now we have

$$\langle F,G\rangle_{\mathscr{Y}_\Gamma} = \langle F_1+F_2, G_1+G_2\rangle = \langle F_1,G_1\rangle + \langle F_1,G_2\rangle + \langle F_2,G_1\rangle + \langle F_2,G_2\rangle.$$

The term $\langle F_1,G_1\rangle$ is

$$\int_{\Gamma\backslash\mathbf{H}} F_1(w)\overline{G_1(w)}\delta_A(y)dA(w)$$

Since $|F_1|^2$ and $|G_1|^2$ are both rapidly decreasing in y, and δ_A cuts off $y \ge A$, this goes to 0 as $A \to \infty$.

The two cross terms have the form

$$\int\int\int \delta_A(y)f(r)g(r')E_A(w,1/2+ir)(y^{1/2+ir} + \phi(1/2+ir)y^{1/2-ir})\,dr\,dr'\,dA(w)$$

which is 0 because A cuts off for large A,where E_A has no constant term.

The final term is

$$\int_{\mathscr{Y}}\int_0^\infty\int_0^\infty f(r)\overline{g(r')}E_A(w,1/2+ir)\overline{E_A(w,1/2+ir')}dr\,dr'\,dA(w).$$

We change the order of integration to move the integral over $\mathscr{Y}$ on the inside, and then apply the Maass-Selberg relation (5.1.1), obtaining

$$\int_0^\infty \int_0^\infty \underbrace{f(r)g(r')\left(\frac{A^{i(r+r')}\phi(1/2+ir')-\phi(1+ir)A^{-i(r+r')}}{i(r+r')}\right)}_{\text{(I)}}$$
$$+\underbrace{f(r)g(r')\left(\frac{A^{i(r-r')}-A^{-i(r-r')}}{i(r-r')}\right)}_{\text{(II)}}+\underbrace{f(r)g(r')\left(\frac{A^{-i(r-r')}(1-\phi(1/2+ir)\overline{\phi}(1/2+ir'))}{i(r-r')}\right)}_{\text{(III)}}dr\,dr'$$

We'll study the limit $A\to\infty$. Note that $\frac{1}{r+r'}$ is bounded on the support of the integrand, since f,g are compactly supported. Hence by the Riemann-Lebesgue Lemma, (I) $\to 0$ as $A\to\infty$.

Since

$$\frac{1-\phi(1/2+ir)\overline{\phi}(1/2+ir')}{i(r-r')}$$

is analytic, Riemann-Lebesgue also implies that (III) $\to 0$ as $A\to\infty$.

Finally (II) is

$$2\int_0^\infty f(r)\frac{\sin(\log A(r-r'))}{r-r'}\,dr \to 2\pi f(r') \text{ as } A\to\infty$$

using that $\int_{-\infty}^\infty \frac{\sin x}{x}\,dx=\pi$, i.e.

$$\frac{1}{2\pi}\int_0^\infty \frac{\sin(\log A\cdot x)}{x}\,dx$$

is an approximate identity as $A\to\infty$. □

Lemma 5.3.1 says that "$E(z,1/2+ir)$ are sort of orthogonal". It is the analogue of Parseval's formula: for

$$\widehat{f}(\xi)=\int f(x)e(-x\xi)dx$$

we have

$$\langle \widehat{f},\widehat{g}\rangle=\langle f,g\rangle.$$

5.4. **Incomplete Eisenstein series.** Let $\psi(y)\in C_c^\infty(\mathbf{R}_{>0})$. We can form the Eisenstein series

$$E(\psi):=\sum_{\gamma\in\Gamma_\infty\backslash\Gamma}\psi(y(\gamma z)).$$

This is a smooth function, in $C_c^\infty(\Gamma\backslash\mathbf{H})$, so it's certainly in $L^2(\Gamma\backslash\mathbf{H})$.

We can determine $E(\psi)$ in terms of Eisenstein series. Consider the Mellin transform

$$\widetilde{\psi}(s)=\int_0^\infty \psi(y)y^{-s}\frac{dy}{y}$$

is entire and decays rapidly in $|t|$ if $s=\sigma+it$ for fixed σ. Then the Mellin inversion formula gives

$$\psi(y)=\frac{1}{2\pi i y}\int_{\mathrm{Re}\, s=\sigma>1}\widetilde{\psi}(s)y^s\,ds.$$

Then we have
$$E(\psi)=\sum\psi(y(\gamma z))=\frac{1}{2\pi i}\int_{\mathrm{Re}\ s=\sigma}\widetilde{\psi}(s)E(z,s)\,ds.$$
Thus we have defined an operator
$$\mathrm{Eis}\colon C_c^{\infty}(\mathbf{R}_{>0})\to\mathrm{Aut}(\Gamma\backslash\mathbf{H})$$
sending $\psi\mapsto E(\psi)$.

Remark 5.4.1. The point of functional analysis is to get good bounds on operators defined on well-behaved functions. This implies that the operators extend automatically by completion.

Now we are going to shift the contour to the line Re $s=1/2$, which is permissible by niceness of $\widetilde{\psi}$. The point is that we residues we pick up are controlled by the results on poles of $E(z,s)$ with Re $s\geq 1/2$ established earlier in Corollary 5.1.2, they are at most simple poles, and are real, with finitely many in $(1/2,1]$ with residues are L^2 eigenfunctions contributing to the discrete spectrum.

We can write
$$E(\psi)=\sum_{\text{residues } s\in(1/2,1]}\widetilde{\psi}(s)+\int_{\mathrm{Re}\ s=1/2}\widetilde{\psi}(\frac{1}{2}+ir)E(\frac{1}{2}+ir)\,dr.$$
So $E(\psi)$ spans a subspace of $L^2(\Gamma\backslash\mathbf{H})$ lying in the continuous spectrum and the space of residues of Eisenstein series. Is there anything else? Let $\mathscr{E}$ be the span of the $E(\psi)$'s in $L^2(\Gamma\backslash\mathbf{H})$, for ψ as above. We will now examine $\mathscr{E}^{\perp}$.

5.5. **Cusp forms.** Let $f\in L^2(\Gamma\backslash\mathbf{H})$ such that $\langle f,E(\psi)\rangle=0$ for all ψ. This condition unravels to
$$\int_{\Gamma\backslash\mathbf{H}}f(z)E(\psi)(z)\,dA(z)=\int_0^{\infty}\int_0^1 f(z)\psi(y)\frac{dx\,dy}{y^2}.$$
If this is to vanish for all ψ, then clearly we must have
$$c(f)=\int_0^1 f(x+iy)\,dx=0\text{ for a.e. }y.$$
Define the space of *cusp forms*
$$L^2_{\mathrm{cusp}}(\Gamma\backslash\mathbf{H}):=\mathscr{E}^{\perp}=\{f\in L^2(\Gamma\backslash\mathbf{H})\colon\int f(x,y)\,dx=0\text{ for a.e. }y\}.$$
This is invariant under Δ and the point-pair algebra $A(\mathbf{H})$.

Let me give a cleaner representation-theoretic description of cusp forms, which leads the way to a generalization for an arbitrary semisimple group.

Let $G=\mathrm{SL}_2(\mathbf{R})$. We define $L^2_{\mathrm{cusp}}(\Gamma\backslash G)$ by the condition
$$\int_{N\cap\Gamma\backslash N}f(ng)\,dn=0\text{ for a.e. }g$$
where $N=\{\begin{pmatrix}1&*\\&1\end{pmatrix}\}$. (In general, you take the unipotent radical of every rational parabolic.)

Clearly $L^2_{\mathrm{cusp}}(\Gamma\backslash G)$ is invariant under the right regular representation. Of course, it could be 0. In fact it often is 0.

Recall that in the proof of the meromorphic continuation, we introduced a cutoff operator

$$\widetilde{k}(z,\zeta)=k(z,\zeta)-\eta_A(y)\int_{\mathbf{R}} k(z,\zeta+t)\,dt.$$

This was a compact operator. Also, $\widetilde{K}|_{L^2_{\text{cusp}}(\Gamma\backslash\mathbf{H})}=K|_{L^2_{\text{cusp}}(\Gamma\backslash\mathbf{H})}$ is both normal and compact. Hence the spectrum of K on cusp forms is *discrete*.

Corollary 5.5.1. *The spectrum of Δ or of $A(\mathbf{H})$ on $L^2_{\text{cusp}}(\Gamma\backslash\mathbf{H})$ is discrete, and there is an orthonormal basis $\phi_1,\dots,\phi_2$ of simultaneous eigenfunctions.*

Finally, we have a spectral decomposition

$$f=\underbrace{\sum_{j=1}^{\ell}\langle f,\psi_i\rangle}_{\text{residual spectrum}}+\underbrace{\int\langle f,E(1/2+ir)\rangle E(z,1/2+ir)\,dr}_{\text{continuous spectrum}}+\underbrace{\sum_j\langle f,\phi_j\rangle\phi_j(z)}_{\text{cuspidal spectrum}}.$$

5.6. **The case of general hyperbolic surfaces.** Let $\Gamma\subset\mathrm{SL}_2(\mathbf{R})$ be a discrete subgroup, such that $X_\Gamma:=\Gamma\backslash\mathbf{H}$ has finite area. Then X_Γ has finitely many cusps, say $\xi_1,\xi_2,\dots,\xi_{n-1},\infty$. For each cusp ξ_j, let y_j be the "y-variable" for the jth cusp. Then we can form an Eisenstein series $E_j(z,s)$ for each j. The expansion of E_i in the jth cusp looks like

$$E_i(z,s)=\delta_{ij}(y^j)^s+\phi_{ij}(s)(y_j)^{1-s}+(\text{higher coefficients}).$$

Let

$$\Phi(s)=(\phi_{ij}(s))_{n\times n}.$$

As before, we have

$$\Phi(s)\Phi(1-s)=\mathrm{Id},$$

with $\Phi(1/2+it)$ being unitary. Letting $\phi(s)=\det\Phi(s)$, we have $|\phi(1/2+it)|=1$. The space of *cusp forms* is defined to be the orthogonal complement of the space of Eisenstein series (coming from all cusps). The analytic continuation is proved as before, using cut-offs and auxiliary equations.

Define the Eisenstein vector

$$\underline{E}(z,s)=\begin{pmatrix}E_1(z,s)\\ \vdots\\ E_n(z,s)\end{pmatrix}.$$

It satisfies the functional equation

$$\underline{E}(z,s)=\Phi(1-s)\underline{E}(z,s).$$

We have a spectral decomposition of $L^2(X_\Gamma)$ (coming from Maass-Selberg), into the pieces

- $L^2_{\text{cusp}}(X_\Gamma)$, which has discrete eigenvalues,
- the residues of of Eisenstein series, which also have discrete spectrum, and
- the continuous spectrum coming from Eisenstein series.

Let $u_1,\dots,$ be the eigenfunctions corresponding to the discrete spectrum. From this spectral decomposition we obtain that any f can be written as

$$f=\sum_j\langle f,u_j\rangle+\int_{-\infty}^{\infty}\langle f,E_j(z,1/2+it)\rangle E_j(z,1/2+it)\,dt.$$

6. Applications

6.1. **The Arthur-Selberg trace formula.** The theory of Eisenstein series was developed for the trace formula. We'll state what it says for finite area hyperbolic surfaces (i.e. the rank 1 case), and hint at some applications.

Let g be an even, smooth function on $\mathbf{R}$ of compact support. Let

$$h(t)=\int_{-\infty}^{\infty} g(x)e^{itx}\,dx$$

Remark 6.1.1. Although $h(t)$ is the Fourier transform of g, we want to emphasize that it plays a very different role from g hence should not be viewed as being in a symmetric position: g is compactly supported on $\mathbf{R}$ while h extends to an entire function on $\mathbf{C}$.

For λ_j in the spectrum, write $\lambda_j = 1/4 + t_j^2$, normalized with $t_j \geq 0$ if $\lambda_j \geq 1/4$, and Re $t_j > 0$ if $0 \leq \lambda_j \leq 1/4$. Consider

$$\sum_{j=0}^{\infty} h(t_j) - \frac{1}{2\pi}\int_{-\infty}^{\infty} h(t)\frac{\phi'}{\phi}(1/2+it)\,dt.$$

Note that we haven't yet proved that $\sum_{j=0}^{\infty} h(t_j)$ makes sense. [15]

Theorem 6.1.2 (Trace formula)**.** *With the notation above, we have*

$$\begin{aligned}
\sum_{j=0}^{\infty} h(t_j) - \frac{1}{2\pi}\int_{-\infty}^{\infty} h(t)\frac{\phi'}{\phi}(1/2+it)\,dt \\
= \frac{A(X_\Gamma)}{4\pi}\int_{-\infty}^{\infty} t\cdot\tanh(\pi t)h(t)\,dt + \\
\sum_{\substack{\{R\}_\Gamma \text{ elliptic} \\ \text{(finite order } m)}} \sum_{k=1}^{m-1} \frac{1}{m\sin(\pi k/m)}\int_{-\infty}^{\infty} \frac{e^{-2\pi t k/m}}{1-e^{-2\pi t}} h(t)\,dt \\
+2\sum_{\{P\}_\Gamma}\sum_{k=1}^{\infty} \frac{\log N(P)}{N(P)^{k/2}-N(P)^{-k/2}} g(k\log N(P)) \\
+ C_{\Gamma,n}g(0) + \frac{1}{2}(n - \operatorname{Tr}\Phi(1/2))h(0) - \frac{n}{\pi}\int_{-\infty}^{\infty} h(t)\frac{\Gamma'}{\Gamma}(1+it)\,dt.
\end{aligned}$$

Here

- *n is still the number of cusps, and*
- *P denotes a hyperbolic conjugacy class, which means it can be diagonalized as*

$$\begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}$$

  with $\lambda > 1$ real. By definition, $N(P) = \lambda^2$.

Remark 6.1.3. The left hand side of the trace formula is the "spectral side", and the right hand side is the "geometric side". You think of the right hand side as being the computable one. In the "simple trace formula" (i.e. the case where Γ is cocompact) you see only the first two terms on the right hand side. The Eisenstein series contributes the second term on the left hand side.

[15] Note that since $\lambda_0 = 0$ is an eigenvalue, $t_0 = i/2$ so we are already uses that h admits an analytic continuation.

You can think of this as a formula as a trace on the discrete spectrum, which is the mysterious part.

Proof sketch. Take a kernel function $k(x,y)$ and form

$$K(x,y)=\sum_{\gamma\in\Gamma}k(\gamma x,y).$$

Set$x=y$ and try to compute the trace. By decomposing the kernel spectrally, you get the left side. By unfolding the integral, you get the right side. □

How can you use the formula? If we make f big and long, then $\widehat{f}$ will be very localized. Localizing h or g gives a formula for the "counting function" on the left or right side.

(1) Let $h_R(t)=h_0(t/R)$ for some h_0. As $R\to\infty$, this localizes g to a delta function at 0. So the trace formula becomes

$$\sum_{t_j\le R}1-\underbrace{\frac{1}{2\pi}\int_{-R}^{R}\frac{\phi'}{\phi}(1/2+it)\,dt}_{\ge 0\text{ for large }R}\sim\frac{A(X_\Gamma)}{4\pi}R^2\sim\frac{A(X_\Gamma)}{4\pi}R^2$$

(We are only counting real t_j since the contribution form the finitely many imaginary t_j becomes negligible).

(2) Swapping the roles of g and h, i.e. considering the family $g_R(t)=g_0(t/R)$, the formula instead count the geodesics by length, yielding the "prime geodesic theorem": defining $\pi(X)=\#\{P:N(P)\le X\}$, i.e. the number of closed geodesics of length at most $\log X$, we have $\pi(X)\sim\frac{X}{\log X}$.

Remark 6.1.4. The trace formula is sort of analogous to Riemann's Explicit formula: for h and g as above, writing the zeroes of $\zeta(s)$ as $1/2+i\gamma_j$, then

$$\sum_j -h(\gamma_j)\sim\int\frac{\Gamma'}{\Gamma}(1/4+it)h(t)\,dt-\sum_p\sum_k\frac{\log p}{p^{k/2}-p^{-k/2}}g(k\log p).$$

Langlands and Shelstad understood that the trace formula could be used to many other purposes, especially to prove functoriality. The philosophy is that you can compute the orbital integrals, and you want to use this to understand the eigenvalues. You want to compare the geometric sides of different trace formula, in order to compare the spectral sides.

6.2. **Existence of cusp forms.** We say that $\Gamma\subset\mathrm{SL}_2(\mathbf{Z})$ is a *congruence subgroup* if $\Gamma\supset\Gamma(N)$. In this case one can generalize the computation of $\phi_\Gamma(s)$, expressing it as a ratio of products of (completed) L-functions[16] of the form

$$\frac{\Lambda(2s-1,\chi)}{\Lambda(2s,\chi)}.$$

Hence

$$\int_{-R}^{R}\frac{\phi_\Gamma'}{\phi_\Gamma}(1/2+it)\,dt\ll R\log R$$

and there are no residues in $(1/2,1)$. Plugging this into the trace formula, one sees that there must be cusp forms.

[16] Everything that we call an "L-function" has an Euler product, a Riemann hypothesis, and has a functional equation $s\longleftrightarrow 1-s$. (Note the normalization of the center.) We *do not* include the archimedean factors.

Corollary 6.2.1. *Congruence subgroups have mostly cuspidal spectrum, meaning that writing the trace formula as* $N_{\text{cusp}}(R)+N_{\text{res}}(R)+M(R)\sim\frac{A(X_\Gamma)}{4\pi}R^2$*, we have already*

$$N_{\text{cusp}}(R)\sim\frac{A(X_\Gamma)}{4\pi}R^2$$

Example 6.2.2. There are $\approx k/12$ cusp forms of weight k on $\mathrm{SL}_2(\mathbf{Z})$, and only one Eisenstein series.

It was originally conjectured by Selberg that the spectrum mostly comes from cusp forms in general, but later this was disproved – it should be a special case of the congruence case.

6.3. **Work of Phillips-Sarnak.** The space of $\Gamma\subset\mathrm{SL}_2(\mathbf{R})$ is Teichmüller space, $T(\Gamma)$. The cotangent space to $T(\Gamma)$ at Γ is canonically isomorphic to the space of holomorphic quadratic differentials. We're going to reformulate this in terms of global coordinates.

Let $X=\Gamma\backslash\mathbf{H}$. This gives a global coordinate z on X.

Definition 6.3.1. A *tensor of weight* (m,n) *on* X is an expression of the form $f(dz)^{\otimes m/2}(d\overline{z})^{\otimes n/2}$ on $X=\mathbf{H}/\Gamma$, with f holomorphic, i.e.

$$f(\gamma z)=(cz+d)^m(\overline{cz+d})^n f(z).$$

A *quadratic differential* is a tensor of weight $(4,0)$.

Example 6.3.2. The function y (imaginary part of z) is a tensor of type $(-1,-1)$.

We consider deforming a line element

$$ds^2=\lambda^2(z)|dz+\mu\overline{dz}|^2.$$

For this to be invariant, μ must be a tensor of weight $(-2,2)$. This is called a *Beltrami differential.* To get our hands on such a thing, let Q be a weight 4 holomorphic cusp form (this is the same thing as a quadratic differential). Then $y^2\overline{Q}(z)$ is a tensor of weight $(-2,2)$, so we can consider the deformation

$$ds^2=\lambda_\epsilon^2(z)|dz+\epsilon s y^2\overline{Q(z)dz}|^2.$$

Definition 6.3.3. The *singular set of* Γ is the subset $\sigma(\Gamma)\subset\mathbf{C}$ defined as follows:

(1) For Re $s\geq 1/2$ and $s\neq 1/2$, $\rho\in\sigma(\Gamma)$ if $\rho(1-\rho)$ is an eigenvalue of Δ, with multiplicity,
(2) For $s=1/2$< there is a special definition involving $\Phi(1/2)$.
(3) For Re $s<1/2$, $\rho\in\sigma(\Gamma)$ if it is a pole of and the multiplicity is the order.

It turns out that the singular set deforms very nicely.

Theorem 6.3.4 (Phillips-Sarnak)**.** *If* Γ_t *is a real-analytic curve in* $T(\Gamma)$*, then* $\sigma(\Gamma_t)$ *varies as an algebroid function of* t *(i.e. the members vary real-analytically except for algebraic singularities when they collide).*

The proof relies on a certain "Phillips operator B", which realizes the singular set as its eigenvalues.

Note that singular elements on the line Re $s=1/2$ can only move to the left. So the first derivative of the real part of $\rho_j=1/2+it_j$ is 0, hence the second derivative is the interesting part.

Theorem 6.3.5 (Phillips-Sarnak, "Fermi Golden rule")**.** *Let* Γ *be a congruence subgroup,* Q *a cuspidal Hecke eigenform of weight* 4*. Then we have*

$$\frac{d^2}{dt^2}\operatorname{Re}(\rho_j(t))|_{t=0}\sim -|L(1/2+it_j,Q\times u_j)|^2.$$

6.4. **The Selberg-Ramanujan conjecture.** Let Γ be a congruence subgroup, with no residues in $(1/2, 1)$.

Conjecture 6.4.1. *The smallest non-zero eigenvalue* $\lambda_1(X_\Gamma)$*, which must correspond to a cusp form, satisfies* $\geq 1/4$.

Remark 6.4.2. If ϕ_j is a cusp form, then we have

$$L_\infty(s, \phi_j) = \pi^{-s}\Gamma\left(\frac{s+it_j}{2}\right)\Gamma\left(\frac{s-it_j}{2}\right).$$

Note that if $t_j = 0$, then $L_\infty(s, \phi_j) = \pi^{-s}\Gamma(s/2)^2$. This coincides with the archimedean factor of an Artin L-function attached to an even Galois representation $\rho : G = \mathrm{Gal}(K/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{C})$. Even Galois representations correspond to Maass forms, and odd Galois representations correspond to modular forms of weight one. Icosahedral Artin representations should correspond to Maass forms with eigenvalue $1/4$, so in particular it should be possible to realize $\lambda_1 = 1/4$.

The fact that the inequality is sharp makes it very difficult to attack using analytic methods. There is a variational formulation that leads to weaker bounds.

Lemma 6.4.3. *Let* Γ *be a congruence subgroup. Then*

$$\lambda_1(X_\Gamma) = \inf_{f \perp constants} \frac{\int_{X_\Gamma} |\Delta_H f|^2 \frac{dx\,dy}{y^2}}{\int_{X_\Gamma} |f|^2 \frac{dx\,dy}{y^2}}.$$

Hence by choosing specific f one can give a lower bound, e.g. for $\mathrm{SL}_2(\mathbf{Z})$ we can prove $\lambda_1 \geq 7$ where the true lower bound appears to be about 90.

6.5. **Rankin-Selberg method.** Let f be a holomorphic cusp form of weight k for $\mathrm{SL}_2(\mathbf{Z})$. By unfolding, we have

$$\int_{X_\Gamma} |f(z)|^2 y^k E(z,s) \frac{dx\,dy}{y^2} = \int_0^\infty \int_0^1 |f(z)|^2 y^{s+k-1} \frac{dx\,dy}{y}. \tag{6.5.1}$$

Writing

$$f(z) = \sum_{n=1}^\infty \lambda_f(n) n^{\frac{k-1}{2}} e(nz)$$

(this is normalized to be symmetric about $s = 1/2$), the integral (6.5.1) becomes

$$\sum_{n=1}^\infty \frac{|\lambda_f(n)|^2}{n^s} \Gamma(s+k-1).$$

Hence we've deduced

$$\sum_{n=1}^\infty \frac{|\lambda_f(n)|^2}{n^s} = \frac{1}{\Gamma(s+k-1)} \int_{X_\Gamma} |f(z)|^2 y^k E(z,s) \frac{dx\,dy}{y^2}.$$

This can be used to show

$$\sum_{n \leq X} |\lambda_f(n)|^2 = cX + O(X^{3/5})$$

which leads to $|\lambda_f(n)| \ll n^{3/10}$. This is an illustration of the Rankin-Selberg method, which was a source of inspiration for Deligne's proof of the Weil Conjectures.

Let's turn to an even more interesting application, namely studying the sum

$$\sum_{n\le X}\lambda_f(n)\lambda_f(n+h).$$

The signs of $\lambda_f(n)$ to should fluctuate randomly, so one hopes for square-root cancellation in this sum, which is equivalent to the Ramanujan-Selberg conjecture. How can we control a sum like this? We replace the Eisenstein series with another function. Instead of averaging y^s over Γ/Γ_∞, we start with the function $y^s e^{-hx}$, which is also Γ_∞-invariant. Then

$$(\Delta+s(1-s))(y^s e(-hx)) = -y^{s+2}4\pi^2h^2e(-hx)$$

Now form the Poincaré series

$$U_h(z,s) := \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} y(\gamma z)^s e(-hx(\gamma z)),$$

which converges for Re $s>1$. Then

$$(\Delta+s(1-s))U_h(z,s) = -4\pi^2h^2U_h(z,s+2). \tag{6.5.2}$$

The left hand side of (6.5.2) is defined for Re $s>1$, but the right hand side is defined for Re $s>-1$. If $(\Delta+s(1-s))$ doesn't kill $U_h(z,s)$ then this can be used to analytically continue $U_h(z,s)$ to the left; this breaks down at λ_1. So if the Selberg-Ramanujan conjecture is true, then the equation (6.5.2) implies that $U_h(z,s)$ is analytic for Re $s>1/2$, leading to square-root cancellation.

We explain a bit more on why the analytic continuation of U_h up to Re $s>1/2$ leads to square-root cancellation of the sum $\sum_{n\le X}\lambda_f(n)\lambda_f(n+h)$. Note that there is a formula

$$\frac{1}{\Gamma(s+k-1)}\int_{X_\Gamma} y^k|f(z)|^2U_h(z,s)\frac{dx\,dy}{y^2} = \sum\frac{\lambda_f(n)\lambda_f(n+h)}{(n+m)^s}\left(\frac{\sqrt{nm}}{m+h}\right)^{k-1}, \tag{6.5.3}$$

which is basically just an unfolding. On the other hand, let $\phi_0,\phi_1,\cdots$ be an orthonormal basis of eigenfunctions of Δ on discrete spectrum of $L^2(X_\Gamma)$, with $\phi_0\equiv 1$. Then the spectral decomposition gives you a formal expression

$$U_h(z,s) = \sum_{j=1}^{\infty}\langle U_h(\cdot,s),\phi_j\rangle\phi_j + \int \text{continuous part}.$$

This expression as itself does not make sense, as U_h is not in L^2 (expected to barely miss being L^2 by Selberg-Ramanujan). However, this can still be used to evaluate an integral of a nice function that is integrated against $U_h(z,s)$, in particular it is applicable to the calculation of the LHS of (6.5.3).

As Γ is a congruence subgroup, the continuous contribution is dominated by the cuspidal contribution. For the cuspidal contribution, we need a general estimate for $\int_{X_\Gamma} y^k|f(z)|^2\phi_j\,dA$ which should achieve a polynomial of λ_j of degree independent of j times an exponential factor that matches with the Gamma value of the LHS of (6.5.3) via Stirling's formula. This estimate basically follows from the following general

Proposition 6.5.1 ([Sar94])**.** *Let P be a polynomial in the eigenfunctions ϕ_k's. Then there are constants A,B depending on P only such that*

$$\int_{X_\Gamma} P(\phi_1,\cdots,\phi_k)(z)\phi_j(z)\,dA(z) \le A(|\lambda_j|+1)^B e^{-\pi\sqrt{\lambda_j}/2}.$$

A similar strategy applies to the rest of the estimate. A full argument can be found in e.g. [Sar01, Appendix A].

6.6. **Approaches to Selberg's eigenvalue conjecture.** First of all, the Selberg's $1/4$ conjecture is not true without congruence assumption, even with arithmeticity. Pick a congruence subgroup Γ such that the first Betti number $h_1(X_\Gamma)$ is positive. As Γ^{ab} has at least one copy of $\mathbb{Z}$, we can pick a one-parameter family of unitary characters $\chi_\theta : \Gamma \to S^1$, where $0 \le \theta < 2\pi$ with $\chi_0 = \mathrm{Id}$.

For a unitary character χ, consider the collection $\mathscr{C}_\chi$ of functions u such that $u(\gamma z) = \chi(\gamma)u(z)$ for $\gamma \in \Gamma$. Let Δ_χ be the Laplacian on these χ-twisted functions. Recalling the variational characterization of the zeroth eigenvalue, we define

$$\lambda_0(\chi) = \inf_{u \in \mathscr{C}_\chi} \frac{\int_{X_\Gamma} |\nabla u|^2 dA}{\int_{X_\Gamma} u^2 dA}.$$

The value $\lambda_0(\chi_\theta)$ is continuous as a function θ. Also, if $\lambda_0(\chi) = 0$, then $\int_{X_\Gamma} |\nabla u|^2 = 0$, so u is a constant, which means $\chi = 1$.

Choose small enough $\varepsilon > 0$ so that there is $\delta > 0$ such that $|\theta| < \delta$ implies $|\lambda_0(\theta)| < \varepsilon$. Choose $|\theta| < \delta$ such that θ is rational. Then χ_θ has finite image, so $\ker \chi_\theta = \Gamma'$ is a finite index subgroup of Γ. The eigenfunction u_0 achieving the infimum $\lambda_0(\chi_\theta)$ is indeed a function on $X_{\Gamma'}$ as $u_0(\gamma z) = \chi_\theta(\gamma)u_0(z) = u_0(z)$ for $\gamma \in \Gamma'$. Thus, u_0 is an eigenfunction of Laplacian on $X_{\Gamma'}$ and $\lambda_1 \ge 1/4$ fails for it.

On the other hand, Selberg himself provided an evidence of his conjecture by proving the following

Theorem 6.6.1 (Selberg)**.** $\lambda_1(X_\Gamma) \ge 3/16$ *for congruence* Γ.

How? An idea is that to work in reverse, namely by using a bound on $\sum_{n \le X} \tau(n)\tau(n+h)$ proved by using some other method, e.g. Kloosterman sum, circle method, etc. This is about the same as counting solutions to $ad - bc = h$ inside a ball, and the circle method does this kind of thing very effectively for quadratic equations of five or more variables. On the other hand, the situation we are in is a four-variable case, so we need to use a Kloosterman sum.

Proof sketch. Consider a Kloosterman sum $S(m,n;c) = \sum^*_{x\overline{x} \equiv 1 (\mathrm{mod}\ c)} e(\frac{mx + n\overline{x}}{c})$. By the Weil bound, we have $|S(m,n;p)| \le 2\sqrt{p}$, so in general $S(m,n;c) \ll c^{1/2+\varepsilon}$, as the number of divisors of c is $\ll c^\varepsilon$. This bound implies that the series $Z(m,n,s) := \sum_c^\infty \frac{S(m,n;c)}{c^{2s}}$ can be holomorphically continued to $\mathrm{Re}\ s > 3/4$. By Goldfeld-Sarnak,

$$\int U_m(z,s)\overline{U_n(z,\overline{s}+2)}dA = \frac{4^{-s-1}n^{-2}\Gamma(2s+1)}{\pi\Gamma(s)\Gamma(s+2)} Z(m,n,s) + R(s),$$

where $R(s)$ is holomorphic in $\mathrm{Re}\ s > 1/2$ with an estimate $R(s) \ll \frac{1}{\mathrm{Re}\ s - 1/2}$ in that region. Thus, any exceptional eigenvalue λ_j with $t_j = \sqrt{1/4 - \lambda_j}$ will contribute to a pole of $Z(m,n,s)$ at $s = 1/2 + t_j$. In particular, any exceptional eigenvalue $< 3/16$ will contribute to a pole of $Z(m,n,s)$ at $s > 3/4$. □

Remark 6.6.2. Using E_8 and functoriality, Kim-Sarnak proved the world record of $\lambda_1 \ge 1/4 - (7/64)^2$. This is in some sense the optimal bound that can be obtained using this kind of idea.

There is another elementary proof of a slightly worse bound using density estimates. For a principal congruence subgroup $\Gamma(q)$, consider a point-pair invariant $K(z,\zeta)$ which is a smooth approximation of characteristic function of $d(z,\zeta) \le R$. Spectrally,

$$K(z,\zeta) = \sum_{\text{discrete}} \hat{k}(t_j)\phi_j(z)\phi_j(\zeta) + \int \text{continuous part}.$$

Since we are in a congruence subgroup case, the discrete spectrum is the main bound, which is $\sim \mathrm{vol}(B(R))/|X_{\Gamma(q)}|$, which comes from the constant function (put $z=i$). This in turn is $\sim R^2 q^{-3}$, as $[\Gamma:\Gamma(q)] \sim q^3$.

This can also be achieved by a bare-hand counting argument. We are counting the number of $\gamma \in \Gamma(q)$ with $\|\gamma\| \le R$. Naively, just from $ad-bc=1$, if we choose a and d and use trivial estimate on the divisor function, we get a slightly worse bound $\sim CR^{2+\varepsilon}q^{-2}$. The correct bound, $\sim R^2 q^{-3}$, can be obtained by a (surprisingyly) simple observation that, for all $\gamma \in \Gamma(q)$, $a+d \equiv 2 (\mathrm{mod}\ q^2)$. So, if we first choose $a+d$ and then a, we get $\sim R^2 q^{-3}$, which is a correct order of magnitude.

This line of thought leads to

Theorem 6.6.3 (Sarnak-Xue for compact case, Gambard for non-compact case)**.** *Let $N(\sigma,q)$ be the number of exceptional eigenvalues smaller than $1/4-\sigma^2$ of $\Gamma(q)$ (of course Selberg's conjecture expects this to be zero). Then $N(\sigma,q) \ll |X_{\Gamma(q)}|^{1-2\sigma+\varepsilon}$*

We end this with yet another approach of Bernstein-Kazhdan using representation theory. As $\Gamma(q) \subset \Gamma(1)$ is a normal subgroup, $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ acts on $X_{\Gamma(q)}$ as isometries. For $\lambda > 0$ an exceptional eigenvalue, the eigenspace V_λ is acted by $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ and is not 1-dimensional (as it is not trivial). The character table of $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ says that a nontrivial representation of $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ has dimension at least $\frac{q-1}{2}$. This gives a bound $\lambda_1 \ge 5/36$.

6.7. **Subconvexity.** Suppose that somehow you are interested in approximating $\zeta(1/2+it)$. The "approximate functional equation" gives

$$\zeta(1/2+it) = \sum_{n \le \sqrt{t}}^{\text{smooth}} n^{-1/2-it} + \gamma(1/2+it) \sum_{n \le \sqrt{t}}^{\text{smooth}} n^{-1/2+it}.$$

This gives a trivial upper bound $|\zeta(1/2+it)| \ll |t|^{1/4+\varepsilon}$. Considering cancellation, Weyl improved this to $|t|^{1/6+\varepsilon}$ by nontrivially controlling $\sum_{n\le N} e(n^k\alpha)$ and using a Taylor expansion of n^{it}.

For a holomorphic modular form f, suppose we want to estimate $L(1/2+it,f)$ (here L-functions do not include archimedean factor). The approximate functional equation in this setting gives

$$L(1/2+it,f) = \sum_{n \le t}^{\text{smooth}} \frac{\lambda_f(n)}{n^{1/2+it}} + \gamma(1/2+it) \sum_{n \le t}^{\text{smooth}} \frac{\lambda_f(n)}{n^{1/2-it}}.$$

The trivial bound is then $L(1/2+it,f) \ll |t|^{1/2+\varepsilon}$, which is called the "convex bound"; in general, for an automorphic representation π, the convex bound is $L(1/2,\pi) \ll C(\pi)^{1/4}$, where $C(\pi)$ is the conductor of π. Any better bound is called a subconvex bound. To use Weyl's idea, we need an estimate of Fourier coefficients of modular form. This can only be done in two ways, either by using a trace formula or by putting a modular form in a family.

Remark 6.7.1. The Riemann hypothesis gives a bound of $C(\pi)^\varepsilon$. This easily follows from the three-lines theorem applied to $\log L(s,\pi)$, which is analytic up to $\mathrm{Re}\, s > 1/2$, assuming the Riemman Hypothesis.

The approximate functional equation can give a bound $\int_T^{2T} |L(1/2+it,f)|^2 dt \ll T\log T$ by expanding and bounding each summand. Note that in this process one needs to take a cut-off test function whose Fourier transform is compactly supported, so that the non-diagonal expansion terms become negligible.

There is a more difficult theorem,

Theorem 6.7.2 (Good)**.** $\int_T^{T+H} |L(1/2+it,f)|^2 dt \ll H t^{\varepsilon}$ *for* $H > T^{2/3}$.

This theorem gives $|L(1/2+it,f)| \ll t^{1/3}$ which is a "Weyl-quality" bound. The proof heavily relies on the spectral theory.

6.8. **Selberg vs. Ramanujan.** What is $L^2(\mathbb{H})$? That $\lambda_1 \geq 1/4$ means that the corresponding radial point-pair invariant barely misses L^2. We will see why this is eqvuialent to the Ramanujan's conjecture at infinity.

Let $G = \mathrm{GL}_2(\mathbb{Q}_v)$, and π be an irreducible unitary representation of G. Let $p(\pi)$ be the infimum of all p such that the matrix coefficient of π are in $L^p(G/Z)$. We say that π is tempered if $p(\pi) = 2$. Adelically, a cusp form is a function f such that $\int_{N(\mathbb{Q})\backslash N(\mathbb{A})} f(ng)dn = 0$ for all unipotent radicals N of all parabolic subgroups P of G. The cuspidal spectrum is discretely decomposed, and by the tensor product theorem, each irreducible representation is a tensor product of local representations. In this setting, the Ramanujan conjecture asserts that any local constituent of automorphic cuspidal representation is tempered. Thus, at infinity this is equivalent to Selberg's conjecture.

The Ramanujan conjecture is proved by Deligne-Serre at GL(2) for cusp forms having infinity type holomorphic discrete series or limits of them. This has been considerably generalized to most of cohomological automorphic representations by a work of many mathematicians.

Remark 6.8.1. The Ramanujan conjecture also has an implication to dynamics. The mixing condition in ergodic theory is actually about the convergence of matrix coefficients, so the Ramanujan conjecture tells you that the mixing rate is the fastest for those other than a constant function.

7. Higher rank cases

There are several notions of "rank" one can consider. Suppose that we have a reductive group $G/\mathbb{Q}$ with a maximal compact subgroup $K \subset G(\mathbb{R})$ and a discrete subgroup $\Gamma \subset G(\mathbb{R})$ (mostly arithmetic in our cases) with finite covolume. Let $S = G(\mathbb{R})/K$ be the symmetric space.

- For a field $F = \mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$, the F-rank of G is the dimension of a maximal F-split torus of G. It is obvious that $\mathrm{rank}_{\mathbb{Q}} G \leq \mathrm{rank}_{\mathbb{R}} G \leq \mathrm{rank}_{\mathbb{C}} G$.
- On the other hand, the Γ-rank is the maximum dimension of a closed, simply connected, totally geodesic flat (simply a flat in Riemannian geometry terms) submanifold in a finite cover of $\Gamma\backslash S$. If Γ is arithmetic, or in other words if $\Gamma \subset G(\mathbb{Q})$ and $G(\mathbb{Z})$ are commensurable, then it turns out that Γ-rank is the same as the $\mathbb{Q}$-rank. In this terms, $\mathbb{R}$-rank is the maximum dimension of a closed simply connected flat of S.

  Another geometric characterization of Γ-rank is the maximum dimension of $\mathbb{R}$-split torus which has a proper orbit in $\Gamma\backslash G(\mathbb{R})$. The Γ-rank is always greater than or equal to the $\mathbb{Q}$-rank, as any maximal $\mathbb{Q}$-split torus has a proper orbit in $G(\mathbb{R})/G(\mathbb{Z})$.

Remark 7.0.1. It is proved by Margulis that a higher $\mathbb{R}$-rank group does not have a non-arithmetic lattice.

For any case of Γ-rank 1 and $\mathbb{R}$-rank 1 (e.g. $G/K = \mathbb{H}^n$, the hyperbolic n-space, with a non-cocompact lattice $\Gamma \subset G$), the proof we have seen above using Fredholm theory and auxiliary equations works exactly the same.

7.1. **Γ-rank 1, higher $\mathbb{R}$-rank cases.** A basic example for Γ-rank 1 and $\mathbb{R}$-rank ≥ 2 case is that of a Hilbert-Blumenthal group.

Let $K=\mathbb{Q}(\sqrt{d})$ for a squarefree integer $d\geq 2$. For $\Gamma=\mathrm{SL}_2(\mathcal{O})$, where $\mathcal{O}$ is the ring of integers of K, we have a natural embedding of $\Gamma\hookrightarrow \mathrm{SL}_2(\mathbb{R})\times\mathrm{SL}_2(\mathbb{R})$ via $\gamma\mapsto(\gamma,\overline{\gamma})$. Through this twisted diagonal embedding, $\Gamma=\mathrm{SL}_2(\mathcal{O})$ acts on $\mathbb{H}\times\mathbb{H}$ discontinuously, and it has a finite volume quotient.

In this case, the symmetric space is $S=\mathbb{H}\times\mathbb{H}$, which has $\mathbb{R}$-rank 2, with a maximal flat given by for example $F=\{(0,y_1)\times(0,y_2)\}$. The ring of invariant differential operators $D(S)$ is generated by two operators, Δ_1 and Δ_2, where Δ_i is the Laplacian in z_i, the complex coordinate of the i-th $\mathbb{H}$-factor of $S=\mathbb{H}\times\mathbb{H}$. In general, $D(S)$ is always a commutative ring generated by $\mathrm{rank}_{\mathbb{R}}\,G$ algebraically independent elements.

The locally symmetric space $\Gamma\backslash\mathbb{H}\times\mathbb{H}$ is a four-dimensional real manifold. Note that $\Gamma\backslash\mathbb{H}\times\mathbb{H}$ is very different from $\Gamma\backslash\mathbb{H}\times\Gamma\backslash\mathbb{H}$. In particular, the projection of the twisted diagonal $\Gamma\leq\mathbb{H}\times\mathbb{H}$ on each factor is dense. This amounts to the fact that the twisted diagonal is an irreducible lattice.

For simplicity, assume $h(\mathcal{O})=1$, which is the same as assuming that there is only one cusp. We take a parabolic subgroup $P=\left\{\begin{pmatrix}\eta & \xi\\ 0 & \eta^{-1}\end{pmatrix}\right\}$, and $\Gamma_\infty:=P\cap\Gamma=\left\{\begin{pmatrix}\eta & \xi\\ 0 & \eta^{-1}\end{pmatrix}\mid \eta\in\mathcal{O}^\times,\xi\in\mathcal{O}\right\}$.

We want to develop a theory of Eisenstein series. To mimic the argument we used in the upper half plane, we want to have a control on an automorphized kernel when approaching to infinity. Thus we need to know how far a Γ-action sends a point when the point is approaching infinity.

Instead of exactly pinning down a fundamental domain, we only need a **Siegel set** that intersects only with finitely many fundamental domains. For example, for $\mathbb{H}$, we can take a box $[-a,a]\times[b,\infty)$. To get a handle on Γ action around infinity, we investigate the set of geodesic rays that are completely contained in the Siegel set. In the upper half plane case, this is a set of vertical rays. We say that two rays are equivalent when the two remains to be within a bounded distance when they are parametrized properly (say by arclength). In this example of $\mathbb{H}$, there is only one equivalence class, we can say the cusp at infinity is “point-like".

Let's get back to our original example of $\Gamma\backslash\mathbb{H}\times\mathbb{H}$. We can take a Siegel set for $\Gamma=\mathrm{SL}_2(\mathcal{O})$ as

$$C=\{(\xi_1,y_1,\xi_2,y_2)\mid|\xi_1|,|\xi_2|\leq c_1,\varepsilon_0^{-1}\leq y_1/y_2\leq\varepsilon_0,y_1,y_2\geq a>0\},$$

for some appropriate c_1,a,ε_0. One can check by hand that this is indeed a Siegel set.

Fix $p_0=((0,1),(0,1))$. If we look at geodesics out of p_0 that remain in C, it is also checkable by hand that this can only happen when the geodesic is in the flat F.

If we change variables to $y_1=e^u, y_2=e^v$, C is characterized by new conditions, $|u-v|\leq c_3$ and $1\leq u+v<\infty$, for some appropriate choice of c_3. Also, the geometry in (u,v) is just that of a Euclidean plane. Thus we know that the boundary $\partial(\mathfrak{F}_\Gamma)$ of a fundamental domain $\mathfrak{F}_\Gamma$ is “point-like".

We now define the Eisenstein series to be

$$E(z,s_1,s_2)=\sum_{\gamma\in\Gamma_\infty\backslash\Gamma}y_1(\gamma z)^{s_1}y_2(\gamma z)^{s_2}.$$

In order that $y_1^{s_1}y_2^{s_2}$ is Γ_∞-invariant, we need $s_1-s_2=2\pi i m/\log\varepsilon_0$, where ε_0 is the fundamental unit and m is an integer. So, the Eisenstein series is really indexed by $E(z,m,s)$, which implies that there are infinitely (countably) many continuous spectra. For each fixed $m\in\mathbb{Z}$, one can meromorphically continue $E(z,m,s)$ as in real rank 1 case. The spectral decomposition we get looks like

$$L^2(\Gamma\backslash\mathbb{H}\times\mathbb{H})=(\text{constant})\oplus(\text{cusp forms})\oplus\bigoplus_{m\in\mathbb{Z}}(E(z,m,s)\text{ part}).$$

7.2. **Higher Γ-rank.** From now on, we change a convention so that Γ is acting on the right.

Let's consider $G = \mathrm{SL}_n(\mathbb{R})$ and $\Gamma = \mathrm{SL}_n(\mathbb{Z})$. Both the real rank and the Γ-rank are $n-1$, and Γ has a finite covolume. In this case, there is only one cusp, but with complicated structure around it.

We write $G = KAN$ for an Iwasawa decomposition, where $K = \mathrm{SO}(n,\mathbb{R})$, A is the group of diagonal matrices with positive entries, and N is the group of unipotent upper triangular matrices. Using the Iwasawa decomposition, we can define a Siegel set in this case.

Definition 7.2.1. For $t, u > 0$, let $C_{t,u} = KA_tN_u$, where A_t is the set of diagonal matrices with $a_{i,i} \le t a_{i+1,i+1}$ and N_u is the set of unipontent upper triangular matrices with off-diagonal entries having absolute values $\le u$.

Proposition 7.2.2 (Minkowski, Siegel)**.** *$C_{\frac{2}{\sqrt{3}},\frac{1}{2}}$ is indeed a Siegel set.*

Proof. It is sufficient to show that the same thing holds with a similar construction for $\mathrm{GL}_n^+(\mathbb{R})$. Let $\Phi(g) = \|ge_1\|$, where $e_1 = \begin{pmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{pmatrix}$ and $\|\cdot\|$ is a Euclidean norm. Then, Φ is a continuous function on G. Also, $\Phi(kan) = \Phi(a)$.

Fix $g \in G$, and consider values of $\Phi(g\gamma) = \|g\gamma e_1\|$ for $\gamma \in \Gamma$. As γe_1 runs over primitive nonzero vectors, Γe_1 is a part of a lattice. Thus, for some $\gamma_0 \in \Gamma$, $\Phi(g\gamma_0)$ attains its minimum.

Lemma 7.2.3. *Let $g = kan$ and $\Phi(g) \le \Phi(g\gamma_0)$ for all $g \in \Gamma$. Then $a_{11} \le 2a_{22}/\sqrt{3}$.*

Proof of lemma. Note that $\Phi(gu) = \Phi(g)$ for $u \in N(\mathbb{Z})$. Thus, we can assume that $|n_{ij}| \le 1/2$. Let $\delta \in \Gamma$ be an element which swaps e_1, e_2 and fixes $e_3, \cdots, e_n$. Then

$$g\delta(e_1) = g(e_2) = kan(e_2) = ka(e_2 + n_{12}e_1) = k(a_{22}e_2 + a_{11}n_{12}e_1).$$

Thus $\|ge_1\|^2 = a_{11}^2 \le a_{22}^2 + a_{11}^2/4$, which gives the conclusion. $\square$

Lemma 7.2.4. *For $x \in G$, the minimum of $\Phi(x\Gamma)$ is attained over $x\Gamma \cap C_{\frac{2}{\sqrt{3}},\frac{1}{2}}$.*

Proof of lemma. Let $y \in x\Gamma$ be an element over which Φ attains its minimum amongst $\Phi(x\Gamma)$. Let $k_y \in K$ be such that $k_y^{-1}y = \begin{pmatrix} a_{11} & * \\ 0 & b \end{pmatrix}$, for some $(n-1)\times(n-1)$ matrix b. Then $\Phi(k_y^{-1}y) \le \Phi(k_y^{-1}\Gamma)$. By induction, there is $z' \in \mathrm{SL}_{n-1}(\mathbb{Z})$ such that $bz' \in C_{\frac{2}{\sqrt{3}},\frac{1}{2}}$. Using the Iwasawa decomposition with similar choice of K', A', N' for $\mathrm{SL}_{n-1}(\mathbb{R})$, we get a decomposition $bz' = k'a'n'$, which yields

$$k_y^{-1}y\begin{pmatrix} 1 & 0 \\ 0 & z' \end{pmatrix} = \begin{pmatrix} a_{11} & * \\ 0 & k'a'n' \end{pmatrix} = k''a''n'',$$

where $n'' = \begin{pmatrix} 1 & 0 \\ 0 & n' \end{pmatrix}$, $k'' = k_y\begin{pmatrix} 1 & 0 \\ 0 & k' \end{pmatrix} \in K$, $a'' = \begin{pmatrix} a_{11} & 0 \\ 0 & a' \end{pmatrix}$. Obviously $n'' \in N_{\frac{1}{2}}$. Also, applying the above lemma, we get $a''_{11} \le \frac{2}{\sqrt{3}}a''_{22}$. Thus $a'' \in A_{\frac{2}{\sqrt{3}}}$. As $y\begin{pmatrix} 1 & 0 \\ 0 & z' \end{pmatrix} = k_yk''a''n'' \in C_{\frac{2}{\sqrt{3}},\frac{1}{2}}$ and $\Phi(y) = \Phi(y\begin{pmatrix} 1 & 0 \\ 0 & z' \end{pmatrix})$, the claim is proved. $\square$

The above lemmas prove that $C_{\frac{2}{\sqrt{3}},\frac{1}{2}}$ is a Siegel set. $\square$

Exercise 7.2.5. Solve the Pell's equation $x^2 - dy^2 = 1$ by using similar arguments on geometry of numbers.

Hint: $\mathrm{SO}(1,1)/\mathrm{SO}(1,1,\mathbb{Z})$ is compact.

Exercise 7.2.6. A Haar measure on $\mathrm{SL}_n(\mathbb{R})$ can be defined to be $dg = e^{-2\rho(\alpha)} dk\, d\alpha\, dn$ with respect to the Iwasawa decomposition $G = KAN$, where $A = \exp(\alpha_1, \cdots, \alpha_n)$ and ρ is half the sum of positive roots, with respect to the standard choice of everything.

Using the Siegel set $C = C_{\frac{2}{\sqrt{3}}, \frac{1}{2}}$ and the Haar measure dg, show that $\mathrm{vol}(\mathrm{SL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{Z})) < \infty$, by showing that $\int_C dg < \infty$.

Exercise 7.2.7. Let $F(x_1, \cdots, x_n)$ be a rational indefinite quadratic form over $\mathbb{R}$. If F is anisotropic, i.e. if $F(x) = 0$ implies $x = 0$, show that $\mathrm{O}_F(\mathbb{R})/\mathrm{O}_F(\mathbb{Z})$ is compact.

Hint: this is some set of lattices $L \subset \mathrm{SL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{Z})$. The Mahler compactness criterion says that a set $\mathscr{L}$ of lattices is pre-compact if and only if there exist $0 < c_1, c_2 < \infty$ such that, for all $L \in \mathscr{L}$,

- $\mathrm{vol}(L) < c_1$,
- $\min_{v \in L \setminus \{0\}} |v| \geq c_2$.

Now let $C = C_{\frac{2}{\sqrt{3}}, \frac{1}{2}}$ be a Siegel set. As before, fix $p_0 \in C$ and look at geodesic rays out of p_0 that are in C. The torus A of diagonal matrices is a totally geodesic subspace, and actually a maximal flat. One can check that a geodesic can remain inside C only if it is contained in this maximal flat.

For example, if $n = 3$, and if we make a substitution $a_i = e^{\alpha_i}$, then the flat, as a set in $\mathbb{R}^2$ with coordinates α_1, α_2, is

$$\{(\alpha_1, \alpha_2) \mid \alpha_1 \geq \alpha_2 + c, \alpha_2 \geq -\alpha_1 - \alpha_2 + c,$$

for some constant c. Thus, geodesic rays coming out of $p_0 = (0,0)$ in this flat are of form $\alpha_1 = \beta \alpha_2$ for a choice of $-1/2 \leq \beta \leq 1$. From this, we know that the cusps form a line.

The key difference between the rays at the boundary (i.e. $\beta = -1/2$ or 1) and those in between is that the stablizers are different; the boundary rays are of form

$$m_1(t) = \begin{pmatrix} e^t & & \\ & e^t & \\ & & e^{-2t} \end{pmatrix}, m_2(t) = \begin{pmatrix} e^t & & \\ & e^{-t/2} & \\ & & e^{-t/2} \end{pmatrix}.$$

Note that these two rays are precisely the rays with repeated eigenvalues.Thus,

$$\mathrm{Stab}(m_1) = P_1 = \begin{pmatrix} * & * & * \\ * & * & * \\ 0 & 0 & * \end{pmatrix}, \mathrm{Stab}(m_2) = P_2 = \begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & * & * \end{pmatrix},$$

whereas

$$\mathrm{Stab}(m) = B = P_1 \cap P_2 \text{ for } m \neq m_1, m_2.$$

This matters a lot for our purpose, because we want to know the blow-up behavior at infinity of the automorphized point-pair invariant $\sum k(\gamma z, w)$. Such behavior should change depending on whether we approach infinity through boundary rays (with larger stabilizer) or other rays (with smaller stabilizer).

We want to define Eisenstein series for various parabolic subgroups. For SL_n, upon the choice of standard Borel, parabolic subgroups correspond to partitions $n = t_1 + \cdots + t_r$. Let the corresponding parabolic of block upper triangular matrices be denoted as $P_{(t_1, \cdots, t_r)}$. It decomposes as MAN where

- M is the set of block diagonal matrices, with each diagonal block a matrix of determinant 1,
- A is the set of block diagonal matrices of determinant 1, with each block a scalar matrix,
- N is the unipotent radical of this parabolic, consisting of appropriate block unipotent matrices.

We will construct Eisenstein series next time.

REFERENCES

[H08] Helgason, Sigurdur. Geometric analysis on symmetric spaces. Second edition. Mathematical Surveys and Monographs, **39**. American Mathematical Society, Providence, RI, 2008. xviii+637 pp. ISBN: 978-0-8218-4530-1

[Sar94] Sarnak, Peter. Integrals of products of eigenfunctions. Int. Math. Res. Notices **6** (1994), 251–260.

[Sar01] Sarnak, Peter. Estimates for Rankin-Selberg L-functions and quantum unique ergodicity. J. Funct. Anal. **184** (2001), no. 2, 419–453.

[Sar03] Sarnak, Peter. Spectra of hyperbolic surfaces. Bull. Amer. Math. Soc. (N.S.) **40** (2003), no. 4, 441–478.

[Sel56] Selberg, A. Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. J. Indian Math. Soc. (N.S.) **20** (1956), 47–87.

[Se73] Serre, J.-P. A course in arithmetic. Translated from the French. Graduate Texts in Mathematics, **7**. Springer-Verlag, New York-Heidelberg, 1973. viii+115 pp.

[Wei87] Weil, André. Prehistory of the zeta-function. Number theory, trace formulas and discrete groups (Oslo, 1987), 1–9, Academic Press, Boston, MA, 1989.

[WW] Whittaker, E. T.; Watson, G. N. A course of modern analysis. An introduction to the general theory of infinite processes and of analytic functions; with an account of the principal transcendental functions. Reprint of the fourth (1927) edition. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1996. vi+608 pp. ISBN: 0-521-58807-3