Ann. Henri Poincaré 99 (9999), 1–65
1424-0637/99000-0, DOI 10.1007/s00023-003-0000

Annales Henri Poincaré

Patterson-Sullivan Distributions and Quantum Ergodicity

Nalini Anantharaman and Steve Zelditch

Abstract. This article gives relations between two types of phase space distributions associated to eigenfunctions ϕ_{ir_j} of the Laplacian on a compact hyperbolic surface X_Γ:

- Wigner distributions $\int_{S^*\mathbf{X}_\Gamma} a \; dW_{ir_j} = \langle Op(a)\phi_{ir_j}, \phi_{ir_j}\rangle_{L^2(\mathbf{X}_\Gamma)}$, which arise in quantum chaos. They are invariant under the wave group.
- Patterson-Sullivan distributions PS_{ir_j}, which are the residues of the dynamical zeta-functions $\mathcal{Z}(s;a) := \sum_\gamma \frac{e^{-sL_\gamma}}{1-e^{-L_\gamma}} \int_{\gamma_0} a$ (where the sum runs over closed geodesics) at the poles $s = \frac{1}{2} + ir_j$. They are invariant under the geodesic flow.

We prove that these distributions (when suitably normalized) are asymptotically equal as $r_j \to \infty$. We also give exact relations between them. This correspondence gives a new relation between classical and quantum dynamics on a hyperbolic surface, and consequently a formulation of quantum ergodicity in terms of classical ergodic theory.

1. Introduction, statement of results.

The purpose of this article is to relate two kinds of phase space distributions which are naturally attached to the eigenfunctions ϕ_{ir_j} of the Laplacian $\triangle$ on a compact hyperbolic surface $\mathbf{X}_\Gamma$. The first kind are the *Wigner distributions* $W_{ir_j} \in \mathcal{D}'(S^*\mathbf{X}_\Gamma)$ (1.1) of quantum mechanics. The second kind are what we call *normalized Patterson-Sullivan distributions* $\widehat{PS}_{ir_j} \in \mathcal{D}'(S^*\mathbf{X}_\Gamma)$ (1.3). In Theorem 1.3, we prove that the Patterson-Sullivan distributions are the residues of classical dynamical zeta functions at poles in the 'critical strip', and therefore have a purely classical definition. Yet in Theorem 1.1, we prove that there exists an 'intertwining

---

Research partially supported by NSF grant #DMS-0302518 and NSF Focussed Research Grant # FRG 0354386.

operator' L_r (1.6) which transforms $\widehat{PS}_{ir_j}$ into W_{ir_j} and which induces an asymptotic equality $W_{ir_j} \sim \widehat{PS}_{ir_j}$ between them. It follows that some of the principal objects and problems of quantum chaos on a compact hyperbolic surface have a purely classical mechanical interpretation. The full nature of the intertwining relation between quantum and classical dynamics will be investigated further in [AZ]. It should generalize to finite volume hyperbolic manifolds of all dimensions, but seems to be a special feature of locally symmetric manifolds related to uniqueness of triple products (invariant trilinear functionals; see [BR, R]).

To state our results, we introduce some notation. We write $G = PSU(1,1) := SU(1,1)/\pm I \equiv PSL(2,\mathbf{R}), K = PSO(2)$ and identify the quotient G/K with the hyperbolic disc $\mathbf{D}$. We let $\Gamma \subset G$ denote a co-compact discrete group and let $\mathbf{X}_\Gamma = \Gamma\backslash\mathbf{D}$ denote the associated hyperbolic surface. By "phase space" we mean the unit cotangent bundle $S^*\mathbf{X}_\Gamma$, which may be identified with the unit tangent bundle $S\mathbf{X}_\Gamma$ and also with the quotient $\Gamma\backslash G$. By a distribution $E \in \mathcal{D}'(Y)$ on a space Y we mean a continuous linear functional on $\mathcal{D}(Y) = C_0^\infty(Y)$. We denote the pairing of distributions E and test functions f by $\langle f, E\rangle_Y$ or $\int_Y f(y)E(dy)$, depending on convenience. We denote by $\lambda_0 = 0 < \lambda_1 \le \lambda_2 ...$ the spectrum of the Laplacian on $\mathbf{X}_\Gamma$, repeated according to multiplicity; with the usual parametrization $\lambda_j = s_j(1-s_j) = \frac{1}{4} + r_j^2$ $(s_j = \frac{1}{2} + ir_j)$, we denote by $\{\phi_{ir_j}\}_{j=0,1,2,...}$ an orthonormal basis of real-valued eigenfunctions: $\triangle\phi_{ir_j} = -\lambda_j\phi_{ir_j}$.

The Wigner distributions (microlocal lifts, microlocal defect measures...) $W_{ir_j} \in \mathcal{D}'(S^*\mathbf{X}_\Gamma)$ are defined by

$$\langle a, W_{ir_j}\rangle = \int_{S^*\mathbf{X}_\Gamma} a(g)W_{ir_j}(dg) := \langle Op(a)\phi_{ir_j}, \phi_{ir_j}\rangle_{L^2(\mathbf{X}_\Gamma)}, \quad a \in C^\infty(S^*\mathbf{X}_\Gamma) \tag{1.1}$$

where $Op(a)$ is a special quantization of a, defined using hyperbolic Fourier analysis (Definition 3.4). The Wigner distribution W_{ir_j} depends quadratically on ϕ_{ir_j}, has mass one in the sense that $\langle \mathbf{I}, W_{ir_j}\rangle = 1$, and has the quantum invariance property

$$\langle U_t^* Op(a)U_t\phi_{ir_j}, \phi_{ir_j}\rangle = \langle Op(a)\phi_{ir_j}, \phi_{ir_j}\rangle, \quad (U_t = \exp(it\sqrt{\Delta})); \tag{1.2}$$

hence by Egorov's theorem W_{ir_j} is asymptotically invariant under the action of the geodesic flow g^t on $S^*\mathbf{X}_\Gamma$, in the large energy limit $r_j \longrightarrow +\infty$. The Wigner distribution W_{ir_j} is one of the principal objects in quantum chaos: it determines the oscillation and concentration of the eigenfunction ϕ_{ir_j} in the classical phase space $S^*\mathbf{X}_\Gamma$ (see §2). One of the main problems in quantum chaos is the *quantum unique ergodicity problem* of determining which geodesic flow invariant probability measures arise as weak* limits of the Wigner distributions (cf. [Lin, RS, W, Sh, SV, Z2] for a few articles on hyperbolic quotients).

The (non-normalized) Patterson-Sullivan distributions $\{PS_{ir_j}\}$ associated to the eigenfunctions $\{\phi_{ir_j}\}$ (cf. Definition 3.3) are defined by the expression

$$PS_{ir_j}(dg) = PS_{ir_j}(db', db, dt) := \frac{T_{ir_j}(db)T_{ir_j}(db')}{|b-b'|^{1+2ir_j}} \otimes |dt|. \tag{1.3}$$

In this definition, T_{ir_j} is the boundary values of ϕ_{ir_j} in the sense of Helgason (cf. Theorem 3.1 or [He, H].) The parameters (b', b) $(b \neq b')$ vary in $B \times B$, where $B = \partial \mathbf{D}$ is the boundary of the hyperbolic disc, and t varies in $\mathbf{R}$; (b', b) parametrize the space of oriented geodesics, t is the time parameter along geodesics, and the three parameters (b', b, t) are used to parametrize the unit tangent bundle $S\mathbf{D}$.

The Patterson-Sullivan distributions PS_{ir_j} are invariant under the geodesic flow (g^t) on $S\mathbf{D}$, i.e.

$$(g^t)_* PS_{ir_j} = PS_{ir_j}. \tag{1.4}$$

The distributions PS_{ir_j} are also Γ-invariant (cf. Proposition 3.3), hence they define geodesic-flow invariant distributions on $S\mathbf{X}_\Gamma$. We also introduce normalized Patterson-Sullivan distributions

$$\widehat{PS}_{ir_j} := \frac{1}{\langle \mathbf{I}, PS_{ir_j} \rangle_{S\mathbf{X}_\Gamma}} \, PS_{ir_j}, \tag{1.5}$$

which satisfy the same normalization condition $\langle \mathbf{I}, \widehat{PS}_{ir_j} \rangle = 1$ as W_{ir_j} on the quotient $S\mathbf{X}_\Gamma$. In Theorem 1.2, it is shown that $\langle \mathbf{I}, PS_{ir_j} \rangle . 2^{(1+2ir_j)} \mu_0(\frac{1}{2} + ir_j) = 1$ where $\mu_0(s) = \frac{\Gamma(\frac{1}{2})\Gamma(s-\frac{1}{2})}{\Gamma(s)}$. Note that the normalizing factor does not depend on the group Γ.

Phase space distributions of this kind were associated to ground state eigenfunctions of certain infinite area hyperbolic surfaces by S. J. Patterson [Pat0, Pat1], and were studied further by D. Sullivan [Su1, Su2] (see also [N]). Ground state Patterson-Sullivan distributions are positive measures, but our analogues for higher eigenfunctions on compact (or finite area) hyperbolic surfaces are not measures. To our knowledge, they have not been studied for higher eigenfunctions before.

Both families (W_{ir_j}) and $(\widehat{PS}_{ir_j})$ are normalized, Γ-invariant bilinear forms in the eigenfunctions ϕ_{ir_j} with values in distributions on $S\mathbf{X}_\Gamma$. But they possess different invariance properties: the former are invariant under the quantum dynamics (the wave group) while the latter are invariant by the classical evolution (the geodesic flow). The motivating problem in this article is to determine how they are related.

The exact relation involves the operator $L_r : C_0^\infty(G) \to C^\infty(G)$ defined by

$$L_r a(g) = \int_{\mathbf{R}} (1 + u^2)^{-(\frac{1}{2} + ir)} a(g n_u) du \tag{1.6}$$

which, we will see, mediates between the classical and quantum pictures. Here, $n_u = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}$ acts on the right as the horocycle flow. We further introduce a cutoff function $\chi \in C_0^\infty(\mathbf{D})$ which is a smooth replacement for the characteristic function of a fundamental domain for Γ (called a 'smooth fundamental domain cutoff', see Definition 3.2).

Theorem 1.1. *For any $a \in C^\infty(\Gamma \backslash G)$ we have the exact formula*

$$\langle Op(a) \phi_{ir_j}, \phi_{ir_j} \rangle_{S\mathbf{X}_\Gamma} = 2^{(1+2ir_j)} \int_{S\mathbf{D}} (L_{r_j} \chi a)(g) PS_{ir_j}(dg),$$

and the asymptotic formula

$$\int_{S\mathbf{X}_\Gamma} a(g) W_{ir_j}(dg) = \int_{S\mathbf{X}_\Gamma} a(g) \widehat{PS}_{ir_j}(dg) + O(r_j^{-1}).$$

It follows that the Wigner distributions are equivalent to the Patterson-Sullivan distributions in the study of quantum ergodicity. The operators L_r in a sense intertwine classical and quantum dynamics (the precise intertwining relation will be investigated in [AZ]). We note that, although the Wigner distributions were defined by using the special hyperbolic pseudodifferential calculus Op, any other choice of Op will produce asymptotically equivalent Wigner distributions and hence Theorem 1.1 is stable under change of quantization.

When a is an automorphic eigenfunction, i.e. a joint eigenfunction of the Casimir operator Ω and the generator W of K, we can evaluate the first expression in Theorem 1.1 to obtain a very concrete relation:

Theorem 1.2. (0) *The normalization of PS_{ir_j} is given by*

$$1 = \langle Op(\mathbf{1}) \phi_{ir_j}, \phi_{ir_j} \rangle = 2^{(1+2ir_j)} \mu_0(\frac{1}{2} + ir_j) \langle \mathbf{1}, PS_{ir_j} \rangle_{\Gamma\backslash G},$$

where $\mu_0(s) = \frac{\Gamma(\frac{1}{2})\Gamma(s-\frac{1}{2})}{\Gamma(s)}$.

More generally, if σ is an eigenfunction of Casimir parameter τ and weight m in the continuous series, we have:

(i)

$$\begin{aligned}\langle Op(\sigma) \phi_{ir_j}, \phi_{ir_j} \rangle &= 2^{(1+2ir_j)} \mu^c_{m,\tau}(\tfrac{1}{2} + ir_j) \left\langle \sigma, PS_{ir_j} \right\rangle_{\Gamma\backslash G} \\ &\quad + 2^{(1+2ir_j)} \mu^{codd}_{m,\tau}(\tfrac{1}{2} + ir_j) \left\langle X_+\sigma, PS_{ir_j} \right\rangle_{\Gamma\backslash G},\end{aligned}$$

where $\mu^c_{m,\tau}$ and $\mu^{codd}_{m,\tau}$ are defined in (5.6)*; X_+ denotes the vector field generating the horocycle flow.*

(ii) *If $\sigma = \psi_m$ is a lowest weight vector in the holomorphic discrete series, we have*

$$\langle Op(\psi_m) \phi_{ir_j}, \phi_{ir_j} \rangle = 2^{(1+2ir_j)} \mu^d_m(\frac{1}{2} + ir_j) \left\langle \psi_m, PS_{ir_j} \right\rangle_{\Gamma\backslash G},$$

*where μ^d_m is defined in (*5.7*).*

These exact formulae are based on the identity (cf. Proposition 6.4),

$$\int_{S\mathbf{D}} (L_{r_j} \chi\sigma) PS_{ir_j}(dg) = \int_{\mathbf{R}} (1+u^2)^{-(\frac{1}{2}+ir_j)} I_{PS_{ir_j}}(\sigma)(u) du, \tag{1.7}$$

where $I_{PS_{ir_j}} : C^\infty(\Gamma\backslash G) \to C(\mathbf{R})$ is the operator defined by

$$I_{PS_{ir_j}}(\sigma)(u) := \int_{\Gamma\backslash G} \sigma(gn_u) PS_{ir_j}(dg). \tag{1.8}$$

When σ is a joint eigenfunction of the Casimir operator Ω and of the generator W of the maximal compact subgroup K, the function $I_{PS_{ir}}(\sigma)(u)$ is a special function of hypergeometric type depending on r and the eigenvalue parameters of σ (cf. §2 for a review of the representation theory of $L^2(\Gamma\backslash G)$). The integral on the right side of (1.7) can then be evaluated to give the explicit formulae of Theorem 1.2.

In our subsequent article [AZ], we give generalizations of Theorems 1.1 and 1.2 to off-diagonal Wigner and Patterson-Sullivan distributions. The correspondence between Wigner and Patterson-Sullivan distributions determines a type of intertwining between classical and quantum mechanics. It is obvious that there cannot exist an intertwining on the L^2 level, since the quantum dynamics has a discrete L^2 spectrum and classical dynamics has a continuous L^2 spectrum, but the correspondence establishes an intertwining on the level of distributions.

Our next result gives a purely classical dynamical interpretation of the Patterson-Sullivan distributions in terms of closed geodesics. Given $a \in C^\infty(S\mathbf{X}_\Gamma)$, we define two closely related dynamical zeta-functions

$$\begin{cases} (i) \quad \mathcal{Z}_2(a,s) = \sum_\gamma \frac{e^{-(s-1)L_\gamma}}{|\sinh(L_\gamma/2)|^2} \left(\int_{\gamma_0} a\right), \\ (ii) \quad \mathcal{Z}(s;a) := \sum_\gamma \frac{e^{-sL_\gamma}}{1-e^{-L_\gamma}} \left(\int_{\gamma_0} a\right), \quad (\Re e\ s > 1) \end{cases} \tag{1.9}$$

where the sum runs over all closed orbits, and γ_0 is the primitive closed orbit traced out by γ. The sum converges absolutely for $\Re e\ s > 1$.

Theorem 1.3. *Let a be a real analytic function on the unit tangent bundle. Then $\mathcal{Z}(s;a)$ and $\mathcal{Z}_2(s;a)$ admit meromorphic extensions to $\mathbf{C}$. The poles in the critical strip $0 < \Re e\ s < 1$, appear at $s = 1/2 + ir$, where as above $1/4 + r^2$ is an eigenvalue of $\triangle$. For each zeta function, the residue is*

$$\sum_{j: r_j^2 = r^2} \langle a, \widehat{PS}_{ir_j} \rangle_{S\mathbf{X}_\Gamma},$$

where $\{\widehat{PS}_{ir_j}\}$ are the normalized Patterson-Sullivan distributions associated to an orthonormal eigenbasis $\{\phi_{ir_j}\}$.

In §7, the thermodynamic formalism is used to prove that $\mathcal{Z}_2(s;a)$ has a meromorphic extension, and we describe its poles and residues in $\Re e\ s > 0$ in terms of "Ruelle resonances". In particular, Patterson-Sullivan distributions arise as the residues. Previously, this formalism has been used to locate the zeros of Selberg's zeta function [Pol]. We use the methods developed by Rugh in [Rugh92] for real-analytic situations. The techniques are based on the Anosov property of the geodesic flow, and apply in variable curvature. However, the relation between Wigner and Patterson-Sullivan distributions is special to constant curvature.

The meromorphic extension of $\mathcal{Z}_2(s;a)$ and the description of its resonances implies the same result for $\mathcal{Z}(s;a)$. But in §9, we give a different kind of proof using representation theory and the generalized Selberg trace formula of [Z]. It seems to us to give a different kind of insight into the meromorphic extension and it can be used to determine residues and poles outside of the critical strip. For the sake of brevity, we only prove it for symbols a which have only finitely many components in the decomposition of $L^2(\Gamma\backslash G)$ into irreducibles. As explained in §9, the extension of the proof to general analytic symbols is related to the estimates on triple products in [BR2, Sa3], and indeed it seems to require non-trivial refinements of them. Like Theorems 1.1- 1.2, the trace formula establishes an exact relation between the Wigner distributions (which appear on the 'spectral side' of the trace formula) and the geodesic periods $\int_\gamma a$ on the 'sum over Γ' side. No such formula can be expected in variable curvature, and the methods are specific to hyperbolic surfaces.

In conclusion, the results of this paper develop to a new level the close relation between classical and quantum dynamics on hyperbolic surfaces. On the level of eigenvalues and lengths of closed geodesics, this close relation is evident from the Selberg trace formula (cf. §8). As is well-known, the Selberg trace formula on a compact hyperbolic manifold is a special case of the general wave trace formula on a compact Riemannian manifold where the leading order approximation is exact. The exactness of this stationary phase formula is somewhat analogous to the exact stationary phase formula of Duistermaat-Heckman for certain oscillatory integrals, but to our knowledge no rigorous link between these exact formulae is known. An alternative explanation of the close relation between classical and quantum dynamics was suggested by V. Guillemin in [G], who made a formal application of the Lefschetz formula to the action of the geodesic flow on a non-elliptic complex. The trace on chains gave the logarithmic derivative of the (Ruelle) zeta function, while the trace on homology gave the spectral side of the Selberg trace formula. For later developments in this direction (by C. Denninger, A. Deitmar, U. Bunke, M. Olbrich and others) we refer to [J].

This paper develops the close relation on the level of eigenfunctions and invariant distributions rather than just eigenvalues and lengths of closed geodesics. As mentioned above, the correspondence between Wigner and Patterson-Sullivan distributions reflects the existence of a kind of intertwining operator between classical and quantum dynamics, which will be investigated further in [AZ]. It is hoped that the intertwining relations will have applications in quantum chaos, e.g. to the question of quantum unique ergodicity. It would also be interesting to relate our constructions to the non-elliptic Lefschetz formulae of [G], to invariant trilinear functionals [BR, R] and to other representation theoretic ones in [SV, W].

2. Background

Hyperbolic surfaces are uniformized by the hyperbolic plane $\mathbf{H}$ or disc $\mathbf{D}$. In the disc model $\mathbf{D} = \{z \in \mathbf{C}, |z| < 1\}$, the hyperbolic metric has the form

$$ds^2 = \frac{4|dz|^2}{(1-|z|^2)^2}.$$

The group of orientation-preserving isometries can be identified with $PSU(1,1)$ acting by Moebius transformations; the stabilizer of 0 is $K \simeq SO(2)$ and thus we will often identify $\mathbf{D}$ with $SU(1,1)/K$. Computations are sometimes simpler in the $\mathbf{H}$ model, where the isometry group is $PSL(2,\mathbf{R})$. We therefore use the general notation G for the isometry group, and G/K for the hyperbolic plane, leaving it to the reader and the context to decide whether $G = PSU(1,1)$ or $G = PSL(2, \mathbf{R})$.

In hyperbolic polar coordinates centered at the origin 0, the Laplacian is the operator

$$\triangle = \frac{\partial^2}{\partial r^2} + \coth r \frac{\partial}{\partial r} + \frac{1}{\sinh^2 r} \frac{\partial^2}{\partial \theta^2}.$$

The distance on $\mathbf{D}$ induced by the Riemannian metric will be denoted $d_{\mathbf{D}}$. We denote the volume form by $dVol(z)$.

Let $\Gamma \subset G$ be a co-compact discrete subgroup, and let us consider the automorphic eigenvalue problem on G/K:

$$\begin{cases} \triangle\phi = -\lambda\phi, \\ \\ \phi(\gamma z) = \phi(z) \text{ for all } \gamma \in \Gamma \text{ and for all } z. \end{cases} \tag{2.1}$$

In other words, we study the eigenfunctions of the Laplacian on the compact surface $\mathbf{X}_\Gamma = \Gamma \backslash\, G\,/K$. Following standard notation (e.g. [V, O]), the eigenvalue can be written in the form $\lambda = \lambda_r = \frac{1}{4} + r^2$ and also $\lambda = \lambda_s = s(1-s)$ where $s = \frac{1}{2} + ir$.

Notational remarks
(i) We denote by $\{\lambda_j = \frac{1}{4} + r_j^2\}$ the set of eigenvalues repeated according to multiplicity, and (in a somewhat abusive manner) we denote a corresponding orthonormal basis of eigenfunctions by $\{\phi_{ir_j}\}$.
(ii) We follow the notational conventions used in [N] and [O], which differ from those used in [H] by a factor 4. We caution that [L, Z] use the latter conventions, and there the parameter s is defined so that $4\lambda = (s-1)(s+1)$ and so that $s = 2ir$.

2.1. Unit tangent bundle and space of geodesics

We denote by $B = \{z \in \mathbf{C}, |z| = 1\}$ the boundary at infinity of $\mathbf{D}$. The unit tangent bundle $S\mathbf{D}$ of the hyperbolic disc $\mathbf{D}$ is by definition the manifold of unit vectors in the tangent bundle $T\mathbf{D}$ with respect to the hyperbolic metric. We may, and will, identify $S\mathbf{D}$ with the unit cosphere bundle $S^*\mathbf{D}$ by means of the metric. We will make a number of further identifications:

- $S\mathbf{D} \equiv PSU(1,1)$. This comes from the fact that $PSU(1,1)$ acts freely and transitively on $S\mathbf{D}$. Similarly, if we work with the upper half plane model $\mathbf{H}$, we have $S\mathbf{H} \equiv PSL(2,\mathbf{R})$. We identify a unit tangent vector (z, v) with a group element g if $g \cdot (i, (0,1)) = (z, v)$. We identify $S\mathbf{D}$, $S\mathbf{H}$, $PSU(1,1)$, and $PSL(2,\mathbf{R})$. In general, we work with the model which simplifies the calculations best. According to a previous remark, $S\mathbf{D}$, $PSU(1,1)$ and $PSL(2,\mathbf{R})$ will often be designated by the letter G.
- $S\mathbf{D} \equiv \mathbf{D} \times B$. Here, we identify $(z, b) \in \mathbf{D} \times B$ with the unit tangent vector (z, v), where $v \in S_z\mathbf{D}$ is the vector tangent to the unique geodesic through z ending at b.

The geodesic flow g^t on $S\mathbf{D}$ is defined by $g^t(z,v) = (\gamma_v(t), \gamma_v'(t))$ where $\gamma_v(t)$ is the unit speed geodesic with initial value (z,v). The space of geodesics is the quotient of $S\mathbf{D}$ by the action of g^t. Each geodesic has a forward endpoint b and a backward endpoint b' in B, hence the space of geodesics of $\mathbf{D}$ may be identified with $B \times B \setminus \Delta$, where Δ denotes the diagonal in $B \times B$: To $(b', b) \in B \times B \setminus \Delta$ there corresponds a unique geodesic $\gamma_{b',b}$ whose forward endpoint at infinity equals b and whose backward endpoint equals b'.

We then have the identification

$$S\mathbf{D} \equiv (B \times B \setminus \Delta) \times \mathbf{R}.$$

The choice of time parameter is defined – for instance – as follows: The point $(b', b, 0)$ is by definition the closest point to 0 on $\gamma_{b',b}$ and (b', b, t) denotes the point t units from $(b, b', 0)$ in signed distance towards b.

2.2. Non-Euclidean Fourier analysis

Following [H], we denote by $\langle z, b\rangle$ the signed distance to 0 of the horocycle through the points $z \in \mathbf{D}$, $b \in B$. Equivalently,

$$e^{\langle z,b\rangle} = \frac{1-|z|^2}{|z-b|^2} = P_{\mathbf{D}}(z,b),$$

where $P_{\mathbf{D}}(z,b)$ is the Poisson kernel of the unit disc. (We caution again that $e^{\langle z,b\rangle}$ is written $e^{2\langle z,b\rangle}$ in [H, Z2]). We denote Lebesgue measure on B by $|db|$, so that the harmonic measure issued from 0 is given by $P_{\mathbf{D}}(z,b)|db|$. A basic identity (cf. [H]) is that

$$\langle g\cdot z, g\cdot b\rangle = \langle z,b\rangle + \langle g\cdot 0, g\cdot b\rangle, \tag{2.2}$$

which implies

$$P_{\mathbf{D}}(gz, gb)\,|d(gb)| = P_{\mathbf{D}}(z,b)\,|db|. \tag{2.3}$$

The functions $e^{(\frac{1}{2}+ir)\langle z,b\rangle}$ are hyperbolic analogues of Euclidean plane waves $e^{i\langle x,\xi\rangle}$ and are called non-Euclidean plane waves in [H]. The non-Euclidean Fourier transform is defined by

$$\mathcal{F}u(r,b) = \int_{\mathbf{D}} e^{(\frac{1}{2}-ir)\langle z,b\rangle} u(z) dVol(z).$$

The hyperbolic Fourier inversion formula is given by

$$u(z) = \int_B \int_{\mathbf{R}} e^{(\frac{1}{2}+ir)\langle z,b\rangle} \mathcal{F}u(r,b) r \tanh(2\pi r) dr |db|.$$

As in [Z3], we define the hyperbolic calculus of pseudo-differential operators $Op(a)$ on $\mathbf{D}$ by

$$Op(a) e^{(\frac{1}{2}+ir)\langle z,b\rangle} = a(z,b,r) e^{(\frac{1}{2}+ir)\langle z,b\rangle}.$$

We assume that the complete symbol a is a polyhomogeneous function of r in the classical sense that

$$a(z,b,r) \sim \sum_{j=0}^{\infty} a_j(z,b) r^{-j+m}$$

for some m (called its order). By asymptotics is meant that

$$a(z,b,r) - \sum_{j=0}^{R} a_j(z,b) r^{-j+m} \in S^{m-R-1}$$

where $\sigma \in S^k$ if $\sup_K (1+r)^{j-k} |D_z^\alpha D_b^\beta D_r^j \sigma(z,b,r)| < +\infty$ for all compact sets K and for all α, β, j.

The non-Euclidean Fourier inversion formula then extends the definition of $Op(a)$ to $C_c^\infty(\mathbf{D})$:

$$Op(a)u(z) = \int_B \int_{\mathbf{R}} a(z,b,r) e^{(\frac{1}{2}+ir)\langle z,b\rangle} \mathcal{F}u(r,b) r \tanh(2\pi r) dr |db|.$$

A key property of Op is that $Op(a)$ commutes with the action of an element $\gamma \in G$ ($T_\gamma u(z) = u(\gamma z)$) if and only if $a(\gamma z, \gamma b, r) = a(z,b,r)$. Γ-equivariant pseudodifferential operators then define operators on the quotient $\mathbf{X}_\Gamma$. This will be seen more clearly when we discuss Helgason's representation formula for eigenfunctions.

2.3. Dynamics and group theory of $G = PSL(2,\mathbf{R})$

We recall the group theoretic point of view towards the geodesic and horocycle flows on $S\mathbf{X}_\Gamma$. As above, it is equivalent to work with $G = PSU(1,1)$ or $G = PSL(2,\mathbf{R})$; we choose the latter. Our notation follows [L, Z], save for the normalization of the metric. The generators of $sl(2,\mathbf{R})$ are denoted by

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad W = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

We denote the associated one parameter subgroups by A, A_-, K. We denote the raising/lowering operators for K-weights by

$$E^+ = H + iV, \quad E^- = H - iV. \tag{2.4}$$

The Casimir operator is then given by $4\ \Omega = H^2 + V^2 - W^2$; on K-invariant functions, the Casimir operator acts as the Laplacian $\triangle$. We also put

$$X_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad X_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

and denote the associated subgroups by N, N_-.

In the identification $S\mathbf{D} \equiv PSL(2,\mathbf{R})$ the geodesic flow is given by the right action of the group of diagonal matrices, A: $g^t(g) = ga_t$ where

$$a_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}.$$

By a slight abuse of notation, we sometimes write a for $\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$. The action of the geodesic flow is closely related to that of the horocycle flow $(h^u)_{u\in\mathbf{R}}$, defined by the right action of N, in other words by $h^u(g) = gn_u$ where

$$n_u = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}.$$

Indeed, the relation $a_t n_u = n_{ue^t} a_t$ shows that the horocyclic trajectories are the stable leaves for the action of the geodesic flow.

The closed orbits of the geodesic flow g^t on $\Gamma\backslash G$ are denoted $\{\gamma\}$ and are in one-to-one correspondence with the conjugacy classes of hyperbolic elements of Γ. We denote by G_γ, respectively Γ_γ, the centralizer of γ in G, respectively Γ. The group Γ_γ is generated by an element γ_0 which is called a primitive hyperbolic geodesic. The length of γ is denoted $L_\gamma > 0$ and means that γ is conjugate, in G, to

$$a_\gamma = \begin{pmatrix} e^{L_\gamma/2} & 0 \\ 0 & e^{-L_\gamma/2} \end{pmatrix}. \tag{2.5}$$

If $\gamma = \gamma_0^k$ where γ_0 is primitive, then we call L_{γ_0} the primitive length of the closed geodesic γ.

2.4. Representation theory of G and spectral theory of $\triangle$

Let us recall some basic facts about the representation theory of $L^2(\Gamma\backslash G)$ in the case where the quotient is compact (cf. [K, L]).

In the compact case, we have the decomposition into irreducibles,

$$L^2(\Gamma\backslash G) = \bigoplus_{j=1}^{S} \mathcal{C}_{ir_j} \oplus \bigoplus_{j=0}^{\infty} \mathcal{P}_{ir_j} \oplus \bigoplus_{m=2,\ m\ even}^{\infty} \mu_\Gamma(m)\mathcal{D}_m^+ \oplus \bigoplus_{m=2, m\ even}^{\infty} \mu_\Gamma(m)\mathcal{D}_m^-,$$

where $\mathcal{C}_{ir_j}$ denotes the complementary series representation, respectively $\mathcal{P}_{ir_j}$ denotes the unitary principal series representation, in which $-\Omega$ equals $s_j(1-s_j) = \frac{1}{4} + r_j^2$. In the complementary series case, $ir_j \in \mathbf{R}$ while in the principal series case $ir_j \in i\mathbf{R}^+$. These continuous series irreducibles are indexed by their K-invariant vectors $\{\phi_{ir_j}\}$, which is assumed to be the given orthonormal basis of $\triangle$-eigenfunctions. Thus, the multiplicity of $\mathcal{P}_{ir_j}$ is the same as the multiplicity of the corresponding eigenvalue of $\triangle$.

Further, $\mathcal{D}_m^\pm$ denotes the holomorphic (respectively anti-holomorphic) discrete series representation with lowest (respectively highest) weight m, and $\mu_\Gamma(m)$ denotes its multiplicity; it depends only on the genus of $\mathbf{X}_\Gamma$. We denote by $\psi_{m,j}$ $(j = 1, \ldots, \mu_\Gamma(m))$ a choice of orthonormal basis of the lowest weight vectors of $\mu_\Gamma(m)\mathcal{D}_m^+$ and write $\mu_\Gamma(m)\mathcal{D}_m^+ = \oplus_{j=1}^{\mu_\Gamma(m)} \mathcal{D}_{m,j}^+$ accordingly.

We will also use the notations $\mathcal{C}_{ir_j}, \mathcal{P}_{ir_j}$ and $\mathcal{D}^{\pm}_{m,j}$ for the orthogonal projection operators of $L^2(\Gamma\backslash G)$ onto these subspaces. Thus, for $f \in L^2$ we write

$$f = \sum_j \mathcal{C}_{ir_j}(f) + \sum_j \mathcal{P}_{ir_j}(f) + \sum_{m,j,\pm} \mathcal{D}^{\pm}_{m,j}(f). \tag{2.6}$$

By an automorphic (τ, m)-eigenfunction, we mean a Γ-invariant joint eigenfunction

$$\begin{cases} \Omega \sigma_{\tau,m} = -(\frac{1}{4} + \tau^2)\sigma_{\tau,m} \\ \\ W\sigma_{\tau,m} = im\sigma_{\tau,m}. \end{cases} \tag{2.7}$$

of the Casimir Ω and the generator W of $K = SO(2)$.

We recall that the principal series $\mathcal{P}_{ir}$ representations of $PSL(2, \mathbf{R})$ are realized on the Hilbert space $L^2(\mathbf{R})$ by the action

$$\mathcal{P}_{ir}\begin{pmatrix} a & b \\ & \\ c & d \end{pmatrix} f(x) = |-bx+d|^{-1-2ir} f\left(\frac{ax-c}{-bx+d}\right).$$

The unique normalized K-invariant vector of $\mathcal{P}_{ir_j}$ is a constant multiple of

$$f_{ir,0}(x) = (1+x^2)^{-(\frac{1}{2}+ir)}.$$

The complementary series representations are realized on $L^2(\mathbf{R}, B)$ with inner product

$$\langle Bf, f\rangle = \int_{\mathbf{R}\times\mathbf{R}} \frac{f(x)\overline{f(y)}}{|x-y|^{1-2u}} dxdy$$

and with action

$$\mathcal{C}_{u}\begin{pmatrix} a & b \\ & \\ c & d \end{pmatrix} f(x) = |-bx+d|^{-1-2u} f\left(\frac{ax-c}{-bx+d}\right).$$

When asymptotics as $|r_j| \to \infty$ are involved, we may ignore the complementary series representations and therefore do not discuss them in detail.

Let $\mathbf{C}_+ = \{z \in \mathbf{C} : \Im z > 0\}$. We recall (see [K], §2.6) that $\mathcal{D}^+_m$ is realized on the Hilbert space

$$\mathcal{H}^+_m = \{f \text{ holomorphic on } \mathbf{C}_+, \quad \int_{\mathbf{C}_+} |f(z)|^2 y^{m-2} dxdy < \infty\}$$

with the action

$$\mathcal{D}^+_{m}\begin{pmatrix} a & b \\ & \\ c & d \end{pmatrix} f(z) = (-bz+d)^{-m} f\left(\frac{az-c}{-bz+d}\right). \tag{2.8}$$

The lowest weight vector of $\mathcal{D}^+_m$ in this realization is $(z+i)^{-m}$.

We note that the K-weights in all irreducibles are even. Lowest weight vectors of $\mathcal{D}_m^+$ correspond to (holomorphic) automorphic forms of weight m for Γ in the classical sense of holomorphic functions on $\mathbf{H}$ satisfying

$$f(\gamma \cdot z) = (cz+d)^m f(z), \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \gamma \in \Gamma.$$

A holomorphic form of weight m defines a holomorphic differential of type $f(z)$ $(dz)^{\frac{m}{2}}$ (cf. [Sa2]). Forms of weight n in $\mathcal{P}_{ir}, \mathcal{C}_u, \mathcal{D}_m^{\pm}$ always correspond to differentials of type $(dz)^{\frac{n}{2}}$. Forms of odd weight do not occur in $L^2(\Gamma \backslash PSL(2, \mathbf{R}))$.

2.5. Time reversibility

Time reversal refers to the involution on the unit cosphere bundle defined by $\iota(x, \xi) = (x, -\xi)$. Under the identification $\Gamma \backslash G \sim S^* X_\Gamma$, the time reversal map takes the form $\Gamma g \to \Gamma g w$ where $w = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ is the Weyl element. For $a \in A$ one has $waw = a^{-1}$.

We say that a distribution is time-reversible if $\iota^* T = T$. The distributions of concern in this article all have the property of time-reversibility, originating in the fact that $\triangle$ is a real operator and hence commutes with complex conjugation. This motivates the decomposition of $\mathcal{P}_{ir} = \mathcal{P}_{ir}^+ \oplus \mathcal{P}_{ir}^-$ into 'even' and 'odd' subspaces.

Proposition 2.1. *We have:*

- *Each principal (or complementary) series irreducible contains a one-dimensional space of A-invariant and time-reversal invariant distributions. In the realization on $L^2(\mathbf{R})$, it is spanned by $\xi_r(x) = |x|^{-(\frac{1}{2}+ir)}$.*
- *There exists a unique (up to scalars) A-invariant time-reversal invariant distribution in $\mathcal{D}_m^+$ when $m \equiv 0 (mod\ 4)$ and there exists no time reversal invariant distribution when $m \equiv 2 (mod\ 4)$. In the realization on $\mathcal{H}_m^+$, it is $z^{-m/2}$. Similarly for $\mathcal{D}_m^-$.*

Proof. **(i) The complementary and principal series**

Each principal (or complementary) series irreducible contains a two-dimensional space of A-invariant distributions. In the model on $L^2(\mathbf{R})$ a basis is given by $x_+^{-(\frac{1}{2}+ir)}, x_-^{-(\frac{1}{2}+ir)}$. Indeed, A invariance is equivalent to

$$e^{-t(\frac{1}{2}+ir)} \xi_{ir}(e^t x) = \xi_{ir}(x).$$

Setting $x = \pm 1$ we find that

$$\xi_{ir}^{\pm}(x) = x_{\pm}^{-(\frac{1}{2}+ir)}$$

are invariant distributions supported on $\mathbf{R}_\pm$.

The time reversal operator is given by

$$\mathcal{P}_{ir} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} f(x) = |x|^{-1-2ir} f(-\frac{1}{x}). \tag{2.9}$$

Hence, time reversal invariance is equivalent to

$$f(-\frac{1}{x}) = |x|^{1+2ir} f(x).$$

Under time reversal

$$\mathcal{P}_{ir}(w) x_+^{-(\frac{1}{2}+ir)} = |x|^{-1-2ir} x_-^{(\frac{1}{2}+ir)} = x_-^{-(\frac{1}{2}+ir)}.$$

Hence the unique time reversal invariant distribution is

$$\xi_{ir} = |x|^{-(\frac{1}{2}+ir)}.$$

(ii) The discrete series

Each holomorphic (or anti-holomorphic) discrete series irreducible $\mathcal{D}_m^{\pm}$ contains a unique (up to scalar multiple) A-invariant distribution $z^{-m/2}$. Indeed, to solve

$$\mathcal{D}_m^+ \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \xi_m^+(z) = e^{mt/2} \xi_m(e^t z) = \xi_m^+(z),$$

we put $z = e^{i\theta}$ and obtain

$$\xi_m^+(re^{i\theta}) = r^{-m/2} \xi_m^+(e^{i\theta}),$$

and the only holomorphic solution is $z^{-m/2}$.

In the holomorphic discrete series, the time reversal operator is given by

$$\mathcal{D}_m^+ \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} f(z) = z^{-m} f(-\frac{1}{z}).$$

We observe that $z^{-m/2}$ is time-reversal invariant when $m \equiv 0 \,(\mathrm{mod}\ 4)$ and is anti-invariant when $m \equiv 2 \,(\mathrm{mod}\ 4)$.

The anti-holomorphic discrete series is similar (by taking complex conjugates).

□

Definition 2.1. We denote the time reversal and geodesic flow invariant distribution in $\mathcal{D}'(\Gamma\backslash G) \cap \mathcal{P}_{ir_j}$ by Ξ_{ir_j}, normalized so that $\langle \phi_{ir_j}, \Xi_{ir_j} \rangle = 1$. We denote by $\Xi_{m,j}^{\pm}$ the time reversal and geodesic flow invariant distribution in $\mathcal{D}'(\Gamma\backslash G) \cap \mathcal{D}_m^{\pm}$, normalized so that $\langle \psi_{m,j}, \Xi_{m,j} \rangle = 1$, where $||\psi_{m,j}|| = 1$. Here, we assume $m \equiv 0\,(4)$.

We now consider the action of A, i.e. the geodesic flow, in each irreducible.

Proposition 2.2. *The right action of A, i.e. the geodesic flow g^t, has two invariant subspaces in each irreducible $\mathcal{C}_{ir}, \mathcal{P}_{ir}$, namely the cyclic subspace generated by the weight zero vector ϕ_{ir}, and that generated by $X_+ \phi_{ir}$. The action of A is irreducible in $\mathcal{D}_m^{\pm}$.*

Proof. In the principal series we have

$$\mathcal{P}_{ir} \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} f(x) = e^{t(\frac{1}{2}+ir)} f(e^t x).$$

The subspaces $L^2(\mathbf{R}_+), L^2(\mathbf{R}_-)$ are invariant, or alternatively the spaces of even and odd functions. The action is irreducible in each subspace: the weight zero vector $(1+x^2)^{-(\frac{1}{2}+ir)}$ generates the former, and its derivative generates the latter.

In the discrete series we have

$$\mathcal{D}_m^+ \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} f(z) = e^{mt/2} f(e^t z).$$

The lowest weight vector is cyclic for the action of A.

□

A nice simplification occurring several times in the paper is that the series $\{X_+\phi_{ir_k}\}$ automatically has zero integrals against a time reversal invariant distribution:

Lemma 2.3. *If $T \in \mathcal{D}'(\Gamma\backslash G)$ is time-reversible, then $\langle X_+\phi_{ir_k}, T\rangle = 0$ for all k.*

Proof. We have

$$\begin{aligned} \langle X_+\phi_{ir}, T\rangle &= \langle X_+\phi_{ir}, \iota^* T\rangle \\ &= \langle \iota^*(X_+\phi_{ir}), T\rangle \\ &= -\langle X_+\phi_{ir}, T\rangle. \end{aligned}$$

□

The following is the main application of the representation theory. By the above normalization, all denominators equal one, but we leave them in to emphasize the normalization.

Proposition 2.4. *Let ν denote a time-reversal invariant and geodesic flow invariant distribution on $\Gamma\backslash G$. Let $f \in C^\infty(\Gamma\backslash G)$. Then:*

$$\begin{aligned} \langle f, \nu\rangle &= \sum_j \frac{\langle \mathcal{P}_{ir_j}(f), \Xi_{ir_j}\rangle}{\langle \phi_{ir_j}, \Xi_{ir_j}\rangle} \langle \phi_{ir_j}, \nu\rangle \\ &+ \sum_{\pm, m=2, m\equiv 0(4)}^{\infty} \sum_{j=1}^{\mu_\Gamma(m)} \frac{\langle \mathcal{D}^\pm_{m,j} f, \Xi^\pm_{m,j}\rangle}{\langle \psi_{m,j}, \Xi^\pm_{m,j}\rangle} \langle \psi_{m,j}, \nu\rangle. \end{aligned}$$

Proof. Since ϕ_{ir} and $X_+\phi_{ir}$ generate $\mathcal{P}_{ir}$ under the action of A, any element f in this space may be expressed in the form $\int_{\mathbf{R}} \tilde{f}_{even}(t)\phi_{ir} \circ g^t dt + \int_{\mathbf{R}} \tilde{f}_{odd}(t) X_+\phi_{ir} \circ g^t dt$.

If we pair with the invariant distribution ν we obtain $\int_{\mathbf{R}} \tilde{f}_{even}(t)dt\ \langle \phi_{ir}, \nu\rangle$. On the other hand, if we pair f with Ξ_{ir} we obtain $\int_{\mathbf{R}} \tilde{f}_{even}(t)dt\ \langle \phi_{ir}, \Xi_{ir}\rangle$. Similarly in the discrete series. The statement follows immediately.

□

To apply the Proposition, we need to understand convergence of the series and hence to have bounds on $\langle \mathcal{P}_{ir}(f), \Xi_{ir}\rangle$ and $\langle \mathcal{D}^{\pm}_{m,j} f, \Xi^{\pm}_{m,j}\rangle$ when the denominator is normalized to equal one. Since the complementary series sum is finite, it is not necessary to analyze these terms. The following proposition shows that the distributions are of order one. Here, we say that a distribution T has order s if $\langle f, T\rangle \leq ||f||_{W^s}$ where $W^s(\Gamma\backslash G)$ is the Sobolev space of functions with s derivatives in L^2. The proposition also controls the dependence of the norms in the Casimir parameters ir, m.

Proposition 2.5. *We have:*

- $|\langle \mathcal{P}_{ir}(f), \Xi_r\rangle| \leq Cr^{-1/2} \; ||\mathcal{P}_{ir}(f)||_{W^1}$;
- $|\langle \mathcal{D}^{+}_{m,j} f, \Xi_{m,j}\rangle| \leq Cm^{-1/2} \; ||\mathcal{D}^{+}_{m,j} f||_{W^1}$;

Proof. We prove the results by conjugating to the models above.

We begin with the continuous series and let

$$\mathcal{U}_{ir} : L^2(\Gamma\backslash G) \to L^2(\mathbf{R})$$

be the unitary intertwining operator from $\mathcal{P}_{ir} \subset L^2(\Gamma\backslash G)$ to its realization in $L^2(\mathbf{R})$. Thus, $\mathcal{U}_{ir}\Xi_{ir} = \xi_{ir}$ up to the normalizing constant.

To determine the normalizing constant, we recall (see [Z], p. 59) that

$$\begin{aligned}\langle \mathcal{U}_{ir}\phi_{ir}, \mathcal{U}_{ir}\Xi_{ir}\rangle &= \int_{\mathbf{R}} (1+x^2)^{-(\frac{1}{2}+ir)} |x|^{-(\frac{1}{2}-ir)} dx \\ &= 2 \int_0^\infty (1+x^2)^{-(\frac{1}{2}+ir)} x^{(\frac{1}{2}+ir)} \frac{dx}{x} \\ &= 2 \int_0^\infty (x^{-1}+x)^{-(\frac{1}{2}+ir)} \frac{dx}{x} \\ &= 2\, B(\tfrac{1}{2}(\tfrac{1}{2}+ir), \tfrac{1}{2}(\tfrac{1}{2}+ir)) := 2\, \frac{\Gamma(\frac{1}{4}+\frac{ir}{2})^2}{\Gamma(\frac{1}{2}+ir)}\end{aligned}$$

Here, $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ is the beta-function. From the asymptotics (cf. [GR] 8.328)

$$\Gamma(x+iy) \sim \sqrt{2\pi} e^{-\frac{\pi}{2}|y|} |y|^{x-\frac{1}{2}} \quad (|y| \to \infty) \tag{2.10}$$

of the Γ-function along vertical lines in $\mathbf{C}$, it follows that

$$(\beta_r)^{-1} := \frac{\Gamma(\frac{1}{4}+\frac{ir}{2})^2}{\Gamma(\frac{1}{2}+ir)} \sim Cr^{-1/2}, \;\; (r \to \infty).$$

Next we consider the order of ξ_{ir} as a distribution in the model. We may break up each function in $L^2(\mathbf{R})$ into its even and odd parts with respect to time reversal invariance, and then we only need to consider $\langle f, \xi_{ir}\rangle$ for a time reversal invariant f. Let $\chi_+(x) \in C_0^\infty(\mathbf{R})$ with $\chi_+ = 1$ for $|x| \leq \frac{1}{2}$ and 0 for $|x| > 2$ and with the property that $\chi_+(x)+\chi_+(\frac{-1}{x}) = 1$. Then $\langle f, \xi_{ir}\rangle = \langle(\chi_+ + \chi_+(\frac{-1}{x}))f, \xi_{ir}\rangle$ and (denoting the time reversal (2.9) operator by T)

$$\begin{aligned}\langle \chi_+(\tfrac{-1}{x})f, \xi_{ir}\rangle &= \langle T(\chi_+(\tfrac{-1}{x})f), T\xi_{ir}\rangle\\ &= \langle \chi_+ f, \xi_{ir}\rangle.\end{aligned}$$

Hence we only need to estimate the χ_+ integral. We write $x^{-1/2+ir} = \frac{1}{-1/2+ir}\frac{d}{dx}$ $x^{1/2+ir}$ and integrate by parts. The result is bounded by $C(1+r)^{-1}(||f||_{L^2} + ||\mathcal{P}_{ir}(X_-)f||_{L^2})$. Here, we use that X_+ is represented by $\frac{d}{dx}$.

It follows that for $f \in C^\infty(\Gamma\backslash G)$,

$$\begin{aligned}|\tfrac{\langle f, \Xi_{ir}\rangle}{\langle \phi_{ir}, \Xi_{ir}\rangle}| = |\langle \mathcal{P}_{ir}(f), \Xi_{ir}\rangle| &= \beta_r |\langle \mathcal{U}_{ir}\mathcal{P}_{ir}(f), \xi_{ir}\rangle|\\ &\leq \beta_r ||\tfrac{d}{dx}\mathcal{U}_{ir}\mathcal{P}_{ir}(f)||_{L^2(\mathbf{R})}\\ &= C\beta_r(1+r)^{-1}||X_-\mathcal{P}_{ir}(f)||_{L^2(\Gamma\backslash G)}\\ &\leq r^{-1/2}\ ||\mathcal{P}_{ir}(f)||_{W^1(\Gamma\backslash G)}.\end{aligned}$$

We now consider the discrete series. The normalizing constant is now

$$\langle \psi_m, \Xi_m^+\rangle = \frac{1}{||(z+i)^{-m}||}\int_{\mathbf{C}_+}(z+i)^{-m}\bar{z}^{-m/2}y^{m-2}dxdy.$$

To calculate the constant, we use the isometry

$$T_m : \mathcal{H}_m^+ \to \mathcal{O}^2(\mathbf{D}, d\nu_m),\ \ T_m f(w) = f\left(-i\frac{w+1}{w-1}\right)\left(\frac{-2i}{w-1}\right)^m,$$

where $\mathcal{O}^2(\mathbf{D}, d\nu_m)$ are the holomorphic functions on the unit disc which are L^2 with respect to the measure $d\nu_m = \frac{4}{4^m}(1-|w|^2)^m\frac{dwd\bar{w}}{(1-|w|^2)^2}$ (cf. [L] IX §3).

We have $T_m\psi_m = 1$. Note that 1 is not normalized to have L^2 norm equal to one. It follows that

$$\langle \psi_m, \Xi_m^+\rangle = \frac{4}{4^m\ ||(z+i)^{-m}||}\int_{\mathbf{D}}\left(-i\frac{w+1}{w-1}\right)^{-m/2}\left(\frac{-2i}{w-1}\right)^m(1-|w|^2)^m\frac{dwd\bar{w}}{(1-|w|^2)^2}.$$

We write $w = re^{i\theta}$ and observe that the angular integral equals an r-independent constant times

$$\int_{S^1}\left(\frac{1+re^{i\theta}}{1-re^{i\theta}}\right)^{-m/2}\left(\frac{2i}{1-re^{i\theta}}\right)^m d\theta = \int_{|z|=1}\left(\frac{1+rz}{1-rz}\right)^{-m/2}\left(\frac{-2i}{1-rz}\right)^m\frac{dz}{z} = 2\pi i(-2i)^m,$$

since $\left(\frac{1+rz}{1-rz}\right)^{-m/2}\left(\frac{-2i}{rz-1}\right)^{m}$ is holomorphic in $|z|\leq 1$ for $r<1$. It follows that

$$\langle\psi_m,\Xi_m^+\rangle = C\frac{2^m}{4^m\ ||(z+i)^{-m}||}\int_{\mathbf{D}}(1-|w|^2)^m\frac{dwd\bar{w}}{(1-|w|^2)^2} = C(m-1)^{-1/2},$$

since the L^2-norm of $T_m\psi_m = 1$ equals $\frac{2}{2^m}\left(\int_{\mathbf{D}}(1-|w|^2)^m\frac{dwd\bar{w}}{(1-|w|^2)^2}\right)^{1/2}$ and $\int_{\mathbf{D}}(1-|w|^2)^m\frac{dwd\bar{w}}{(1-|w|^2)^2}$ equals $\frac{1}{m-1}$.

We then need to estimate

$$\langle\Xi_m^+,\bar{f}\rangle = \int_{\mathbf{C}_+} f(z)z^{-m/2}y^{m-2}dxdy.$$

As above, we let χ_+ be a radial function with compact support in $\mathbf{R}_+$ and with $\chi_+(z)+\chi_+(\frac{-1}{z})\equiv 1$. By unitary of time reversal, we again have

$$\langle\chi_+(\tfrac{-1}{z})\Xi_m^+,\bar{f}\rangle \quad = \quad \langle\chi_+\Xi_m^+,\bar{f}\rangle,$$

and thus it suffices to estimate the χ_+ integral. We note that for $m>2$, $z^{-m/2} = \frac{1}{1-m/2}\frac{d}{dx}z^{-m/2+1}$ and that $z^{-m/2+1}\in L^2(|z|<1, y^m\frac{dxdy}{y^2})$. The operator $\frac{d}{dx} = \mathcal{D}_m^+(X_-)$ is skew symmetric with respect to the inner product. Partial integration gives the bound $\frac{1}{1-m/2}||f'||_{L^2}$, hence after normalizing Ξ_m^+ we have

$$\left|\frac{\langle f,\Xi_m^+\rangle}{\langle\psi_m,\Xi_m^+\rangle}\right| \leq Cm^{-1/2}(||f||_{L^2}+||\mathcal{D}_m^+(X_+)f||_{L^2}).$$

□

Remark 2.1. The paper [A-P] studies related estimates in the discrete series from a different point of view.

3. Patterson-Sullivan distributions and microlocal lifts

3.1. Patterson-Sullivan distributions

Let us first recall Helgason's fundamental result about eigenfunctions of the Laplacian on $\mathbf{D}$. In the following theorem, ϕ is any solution of $\triangle\phi = -\lambda\phi$ ($\lambda = \frac{1}{4}+r^2$ where $\lambda, r\in\mathbf{C}$). The function ϕ, defined on $\mathbf{D}$, is not necessarily automorphic. One says that ϕ has exponential growth if there exists $C>0$ such that $|\phi(z)|\leq Ce^{Cd_{\mathbf{D}}(0,z)}$ for all z.

Theorem 3.1. *(*[H]*, Theorems 4.3 and 4.29; see also* [He]*) Let ϕ_{ir} be an eigenfunction with exponential growth, for the eigenvalue $\lambda = \frac{1}{4}+r^2\in\mathbf{C}$. Then there exists a distribution $T_{ir,\phi_{ir}}\in\mathcal{D}'(B)$ such that*

$$\phi_{ir}(z) = \int_B e^{(\frac{1}{2}+ir)\langle z,b\rangle}T_{ir,\phi_{ir}}(db),$$

for all $z\in\mathbf{D}$. The distribution is unique if $\frac{1}{2}+ir\neq 0,-1,-2,\cdots$.

The theorem extends the classical representation theorem for bounded harmonic functions to the case of arbitrary eigenvalues. Note that the kernel $e^{(\frac{1}{2}+ir)\langle z,b\rangle}$ that appears in the representation theorem for eigenfunctions for the eigenvalue λ_r is the generalized Poisson kernel, $P_{\mathbf{D}}^{(\frac{1}{2}+ir)}(z,b)$. The distribution $T_{ir,\phi_{ir}}$ is called the boundary value of ϕ_{ir} and may be obtained from ϕ_{ir} in several explicit ways. One is to expand the eigenfunction into the "Fourier series",

$$\phi_{ir}(z) = \sum_{n\in\mathbf{Z}} a_n \Phi_{r,n}(z), \tag{3.1}$$

in the disc model in terms of the generalized spherical functions $\Phi_{r,n}$ defined by ([H], Theorem 4.16)

$$e^{(\frac{1}{2}+ir)\langle z,b\rangle} = \sum_{n\in\mathbf{Z}} \Phi_{r,n}(z) b^n, \quad b \in B. \tag{3.2}$$

Then (cf. [H], p. 113)

$$T_{ir,\phi_{ir}}(db) = \sum_{n\in\mathbf{Z}} a_n b^n |db|. \tag{3.3}$$

A second way is that, at least when $\Re(ir) > 0$, the boundary value is given by the limit ([H], Theorem 4.27)

$$\lim_{d(0,z)\to\infty} e^{(\frac{1}{2}+ir)d(0,z)} \phi_{ir}(z) = c(ir) T_{ir,\phi_{ir}},$$

where c is the Harish-Chandra c-function and $d(0,z)$ is the hyperbolic distance.

We note that $\lambda_r = s(1-s)$ corresponds to both $s = \frac{1}{2} + ir$ and $1 - s = \frac{1}{2} - ir$. Except when $ir = 0$, the two choices of s give a distinct boundary value and Poisson representation formula. This explains why the notation $T_{ir,\phi_{ir}}$ for boundary values includes both ir and ϕ_{ir}. The irreducible representations corresponding to the pair of parameters are equivalent, and the intertwining operator between them intertwines the two boundary values [Schm]. The map taking one boundary value to the other may also be viewed as a scattering operator (cf. [Ag]). In Theorem 1.3, the Patterson-Sullivan residue corresponding to $\Re e(ir) \geq 0$ is constructed from the boundary value with $\Re e(ir) \geq 0$, while the residue with $\Re e(ir) < 0$ corresponds to the other boundary value. Since the boundary values are essentially equivalent, we generally assume for simplicity of exposition that $\Re e(ir) \geq 0$.

For a fixed orthonormal basis $\{\phi_{ir_j}\}$ we denote $T_{ir_j,\phi_{ir_j}}$ with $\Re e(ir_j) > 0$ more simply by T_{ir_j}. As observed in [Z2], when ϕ_{ir_j} is a Γ-invariant eigenfunction, the boundary values $T_{ir_j}(db)$ have the following invariance property:

$$\begin{aligned} \phi_{ir_j}(\gamma z) = \phi_{ir_j}(z) \quad &\Longrightarrow e^{(\frac{1}{2}+ir_j)\langle\gamma z,\gamma b\rangle} T_{ir_j}(d\gamma b) = e^{(\frac{1}{2}+ir_j)\langle z,b\rangle} T_{ir_j}(db) \\ &\Longrightarrow T_{ir_j}(d\gamma b) = e^{-(\frac{1}{2}+ir_j)\langle\gamma\cdot 0,\gamma\cdot b\rangle} T_{ir_j}(db) \end{aligned} \tag{3.4}$$

This follows from the uniqueness of the Helgason representation (3.1) and by the identities (2.2)-(2.3). Hence the distribution $e_{ir_j} \in \mathcal{D}'(PSL(2,\mathbf{R}))$ defined by

$$\langle f, e_{ir} \rangle_{PSL(2,\mathbf{R})} = \int_{\mathbf{D}\times B} e^{(\frac{1}{2}+ir_j)\langle z,b\rangle} f(z,b) T_{ir_j}(db) dVol(z) \tag{3.5}$$

is Γ-invariant, as well as horocyclic-invariant. Seen as a distribution on the quotient $\Gamma\backslash PSL(2,\mathbf{R})$, e_{ir} may be expanded in a K-Fourier series,

$$e_{ir_j} = \sum_{n\in\mathbf{Z}} \phi_{ir_j,n},$$

and it is easily seen (cf. [Z2]) that $\phi_{ir_j,0} = \phi_{ir_j}$ and that $\phi_{ir_j,n}$ is obtained by applying the nth normalized raising or lowering operator (Maass operator) to ϕ_{ir_j}. More precisely, one applies $(E^{\pm})^n$ (2.4) and multiplies by the normalizing factor $\beta_{2ir_j,n} = \frac{1}{(2ir_j+1\pm 2n)\cdots(2ir_j+1\pm 2)}$. The regularity of these distributions was recently studied in [FF, Co].

At $z = 0$, the K-Fourier series and B-Fourier series coincide and we get

$$T_{ir_j}(db) = \sum_{n\in\mathbf{Z}} \beta_{s,\pm n} \left((E^{\pm})^n \phi_{ir_j}(0)\right) b^n db. \tag{3.6}$$

This gives a third way of obtaining the boundary values from ϕ_{ir_j}.

We will only need some crude estimates on the regularity of the distributions T_{ir_j}. Rather than estimating the regularity of $T_{ir_j}(db)$ using (3.6), which would take us too far afield, we will quote some estimates of Otal [O] which suffice (and indeed are better than necessary) for our applications. Roughly, they say that $T_{ir_j}(db)$ is the derivative of a Hölder continuous function F_{ir_j}. Since its zeroth Fourier coefficient is non-zero, $T_{ir_j}(db)$ is not literally the derivative of a periodic function, but it is the derivative of a function F_{ir_j} on $\mathbf{R}$ satisfying $F_{ir_j}(\theta+2\pi) = F_{ir_j}(\theta)+C_j$ for all $\theta \in \mathbf{R}$. We follow Otal in calling such a function 2π-periodic.

For $0 \leq \delta \leq 1$ we say that a 2π-periodic function $F : \mathbf{R} \to \mathbf{C}$ is δ-Hölder if $|F(\theta) - F(\theta')| \leq C|\theta - \theta'|^\delta$. The smallest constant is denoted $||F||_\delta$ and Λ_δ denotes the Banach space of δ-Hölder functions, up to additive constants.

Theorem 3.2. ([O] *Proposition 4) Suppose that* $s = \frac{1}{2} + ir$ *with* $\Re s \geq 0$*, and that* ϕ *is an eigenfunction of eigenvalue* $s(1-s)$ *satisfying* $||\nabla\phi||_\infty < \infty$*. Then its Helgason boundary value* $T_{s,\phi}$ *is the derivative of a* $\Re s$*-Hölder function.*

In our case, the theorem says that T_{ir_j} is the derivative of a Hölder function, of Hölder exponent $\frac{1}{2}$ if $\lambda_j \geq \frac{1}{4}$. Otal's proof also shows that the Hölder norm is bounded by a power of r_j. Related results can be found in [BR, C, MS, FF, Co].

We now introduce a "Patterson-Sullivan" distribution associated to each automorphic eigenfunction. Recall that we denote by $\lambda_0 = 0 < \lambda_1 \leq ...$ the spectrum of the Laplacian on $\mathbf{X}_\Gamma$ ($\lambda_j = \frac{1}{4} + r_j^2$), and by (ϕ_{ir_j}) a given orthonormal basis of eigenfunctions whose boundary values are denoted (T_{ir_j}).

Remark 3.1. We assume that these eigenfunctions are real to obtain time reversal invariant distributions. Aside from that, our results are valid for complex eigenfunctions with slight modifications. As mentioned above, we also assume for simplicity that $\Re ir > 0$. The case $\Re ir <$ is similar.

Definition 3.1. The Patterson-Sullivan distribution associated to a real eigenfunction ϕ_{ir_j} is the distribution on $B \times B \setminus \Delta$ defined by

$$ps_{ir_j}(db', db) := \frac{T_{ir_j}(db)T_{ir_j}(db')}{|b - b'|^{1+2ir_j}}$$

If ϕ_{ir_j} is Γ-automorphic, it is easy to check that ps_{ir_j} is invariant under the diagonal action of Γ:

Proposition 3.3. *Suppose that ϕ_{ir_j} is Γ-invariant, and let T_{ir_j} denote its radial boundary values. Then the distribution on $B \times B \setminus \Delta$ defined by*

$$ps_{ir_j}(db', db) := \frac{T_{ir_j}(db)T_{ir_j}(db')}{|b - b'|^{1+2ir_j}}$$

is Γ-invariant and time reversal invariant.

Proof. It follows from (3.4) that

$$T_{ir_j}(d\gamma b)T_{ir_j}(d\gamma b') = e^{-(\frac{1}{2}+ir_j)\langle \gamma\cdot 0, \gamma\cdot b\rangle} e^{-(\frac{1}{2}+ir_j)\langle \gamma\cdot 0, \gamma\cdot b'\rangle} T_{ir_j}(db)T_{ir_j}(db'). \quad (3.7)$$

We will also need the following identities (cf. [N] (1.3.2)):

$$\begin{aligned} |\gamma(x) - \gamma(y)| &= |\gamma'(x)|^{\frac{1}{2}}|\gamma'(y)|^{\frac{1}{2}}|x - y| \\ 1 - |\gamma(x)|^2 &= |\gamma'(x)|(1 - |x|^2). \end{aligned} \quad (3.8)$$

for every $x, y \in \mathbf{D} \cup B$, $\gamma \in \Gamma$. Hence for $b \in B$ and $\gamma \in \Gamma$, we have

$$|\gamma(0) - \gamma(b)|^2 = |\gamma'(b)|(1 - |\gamma(0)|^2). \quad (3.9)$$

Furthermore,

$$|\gamma b - \gamma b'|^2 = e^{-[\langle \gamma\cdot 0, \gamma\cdot b\rangle + \langle \gamma\cdot 0, \gamma\cdot b'\rangle]}|b - b'|^2. \quad (3.10)$$

Raising (3.10) to the power $\frac{1}{2} + ir_j$, taking the ratio with (3.7) and simplifying completes the proof of Γ-invariance.

Time-reversal invariance is invariance under $b \iff b'$, which is obvious from the formula. □

We now construct from the distribution ps_{ir_j} a geodesic flow invariant distribution on $S\mathbf{D}$ as follows. As reviewed in §2, the unit tangent bundle $S\mathbf{D}$ can be identified with $(B \times B \setminus \Delta) \times \mathbf{R}$: the set $B \times B \setminus \Delta$ represents the set of oriented geodesics, and $\mathbf{R}$ gives the time parameter along geodesics. We then define the Radon transform:

$$\mathcal{R} : C_0(S\mathbf{D}) \to C_0(B \times B \setminus \Delta), \quad \text{by } \mathcal{R}f(b', b) = \int_{\gamma_{b',b}} f dt. \quad (3.11)$$

Further, we need to define special cutoffs which have the property that

$$\int_{\mathcal{D}} f dVol(z) = \int_{\mathbf{D}} \chi f dVol(z) \quad (3.12)$$

for any $f \in C(\Gamma\backslash\mathbf{D})$, where $\mathcal{D}$ is a fundamental domain for Γ in $\mathcal{D}$. In other words, χ is a smooth replacement for the characteristic function of $\mathcal{D}$.

Definition 3.2. We say that $\chi \in C_0^\infty(\mathbf{D})$ is a smooth fundamental domain cutoff if

$$\sum_{\gamma\in\Gamma} \chi(\gamma z) = 1.$$

We then make the basic definitions:

Definition 3.3.
1. On $S\mathbf{D}$ we define the Patterson-Sullivan distribution $PS_{ir_j} \in \mathcal{D}'(S\mathbf{D})$ by:
$$PS_{ir_j}(db', db, dt) = ps_{ir_j}(db', db)|dt|$$
in the sense that
$$\langle a, PS_{ir_j}\rangle_{S\mathbf{D}} = \int_{B\times B\backslash\Delta} (\mathcal{R}a)(b', b) ps_{ir_j}(db', db).$$
2. On the quotient $S\mathbf{X}_\Gamma = \Gamma\backslash S\mathbf{D} = \Gamma\backslash PSU(1,1)$, we define the Patterson-Sullivan distributions $PS_{ir_j} \in \mathcal{D}'(S\mathbf{X}_\Gamma)$ by
$$\langle a, PS_{ir_j}\rangle_{S\mathbf{X}_\Gamma} = \langle \chi a, PS_{ir_j}\rangle_{S\mathbf{D}} = \int_{B\times B\backslash\Delta} \mathcal{R}(\chi a)(b', b) ps_{ir_j}(db', db),$$
where χ is a smooth fundamental domain cutoff.
3. As in the introduction (see 1.5), we also define normalized Patterson-Sullivan distributions by
$$\widehat{PS}_{ir_j} := \frac{1}{\langle \mathbf{I}, PS_{ir_j}\rangle_{S\mathbf{X}_\Gamma}}\ PS_{ir_j}.$$

The following proposition is obvious from the definition, but important:

Proposition 3.4. *PS_{ir_j} is a geodesic flow invariant and Γ-invariant distribution on $S\mathbf{D} = \mathbf{D}\times B$; in the quotient, PS_{ir_j} is geodesic flow invariant on $S\mathbf{X}_\Gamma$.*

The geodesic flow invariance of PS_{ir_j} on $S\mathbf{D}$ is trivial; on the quotient $S\mathbf{X}_\Gamma$ it is also easy, and results from the following principle:

Lemma 3.5. *Let $T \in \mathcal{D}'(S\mathbf{D})$ be a Γ-invariant distribution. Let a be a Γ-invariant smooth function on $S\mathbf{D}$. Then, for any $a_1, a_2 \in \mathcal{D}(S\mathbf{D})$ such that $\sum_{\gamma\in\Gamma} a_i(\gamma.(z,b)) = a(z,b)$ $(i = 1,2)$ we have*

$$\langle a_1, T\rangle_{S\mathbf{D}} = \langle a_2, T\rangle_{S\mathbf{D}}$$

Proof. Let χ be a function on $C_0^\infty(\mathbf{D} \times B)$ such that $\sum_{\gamma\in\Gamma} \chi(\gamma.(z,b)) \equiv 1$ (in general, we choose χ to be independent of b). For any such χ we have

$$\begin{aligned} \langle a_i, T\rangle_{S\mathbf{D}} &= \int_{S\mathbf{D}} \{\sum_{\gamma\in\Gamma} \chi(\gamma(z,b))\} a_i(z,b) T(dz,db) \\ &= \int_{S\mathbf{D}} \sum_{\gamma\in\Gamma} \chi(z,b) a_i(\gamma(z,b)) T(dz,db) \\ &= \int_{S\mathbf{D}} \chi(z,b) a(z,b) T(dz,db). \end{aligned}$$

□

If we look at the expression

$$\langle a, PS_{ir_j}\rangle_{S\mathbf{D}} = \int |b-b'|^{-1-2ir_j} \mathcal{R}(a) T_{ir_j}(db) T_{ir_j}(db'), \tag{3.13}$$

and apply Otal's theorem saying that $T_{ir_j} = F'_{ir_j}$ for some Hölder function F_{ir_j}, we easily derive:

For any $a \in C^\infty(S\mathbf{D})$ we have

$$|\langle a, PS_{ir_j}\rangle_{S\mathbf{D}}| \leq ||F_{ir_j}||^2_{L^\infty(B)} \cdot ||\frac{\partial^2}{\partial b \partial b'} |b-b'|^{-1-2ir_j} \mathcal{R}(a)||_{L^\infty(B\times B\setminus\Delta)}$$

provided the left-hand side is well defined. A priori, the right side may be infinite.

For future reference, we state a sufficient condition to obtain a non-trivial estimate:

Proposition 3.6. *Assume that* $|b-b'|^{-1-2ir_j}\mathcal{R}(a) \in C^2(B\times B)$. *Then* $\langle a, PS_{ir_j}\rangle_{S\mathbf{D}}$ *is well defined, and*

$$|\langle a, PS_{ir_j}\rangle_{S\mathbf{D}}| \leq ||F_{ir_j}||^2_{L^\infty(B)} \cdot ||\frac{\partial^2}{\partial b \partial b'} |b-b'|^{-1-2ir_j} \mathcal{R}(a)||_{L^\infty(B\times B\setminus\Delta)}.$$

A simple example where the condition holds is where $a \in C^2_c(S\mathbf{D})$. In that case, there exist $C > 0$ and $K > 0$ such that:

$$|\langle a, PS_{ir_j}\rangle_{S\mathbf{D}}| \leq C(1+|r_j|)^K ||a||_{C^2} \tag{3.14}$$

for all j. If $a \in C^2(S\mathbf{X}_\Gamma)$, $|\langle a, PS_{ir_j}\rangle_{S\mathbf{X}_\Gamma}| \leq C(1+|r_j|)^K ||a||_{C^2}$ for all j.

3.2. Microlocal lift and Wigner distributions

We now give a precise definition of the matrix elements $\langle Op(a)\phi_{ir_j}, \phi_{ir_j}\rangle$ and hence of the Wigner distributions. When a is a Γ-invariant function on $S\mathbf{D}$, then in the non-Euclidean calculus §2.2 we have

$$Op(a)\phi_{ir_j} := \int_B a(z,b) e^{(\frac{1}{2}+ir_j)\langle z,b\rangle} T_{ir_j}(db). \tag{3.15}$$

Definition 3.4. The Wigner measure of ϕ_{ir_j} is the distribution W_{ir_j} on $S\mathbf{X}_\Gamma = \Gamma\backslash S\mathbf{D}$ defined by:

$$\int_{S\mathbf{X}_\Gamma} a(g) W_{ir_j}(dg) := \langle Op(a)\phi_{ir_j}, \phi_{ir_j}\rangle_{L^2(\mathbf{X}_\Gamma)},$$

where $Op(a)$ is given by (3.15).

To see that W_{ir_j} is a distribution of finite order, we note that $\langle Op(a)\phi_{ir_j}, \phi_{ir_j}\rangle_{L^2(\mathbf{X}_\Gamma)}$ is bounded by the operator norm of $||Op(a)||$ and hence by a C^k norm of a. In fact, Otal's regularity theorem shows that it is of order 1 at most.

We further note that W_{ir_j} is quantum time reversible in the sense that $\langle COp(a)C\phi_{ir_j}, \phi_{ir_j}\rangle = \langle COp(a)C\phi_{ir_j}, \phi_{ir_j}\rangle$, where $Cf = \bar{f}$ is the operator of complex conjugation. Clearly, $COp(a)C = Op(\mathcal{C}a)$ where $\mathcal{C}a(z, b, \lambda) = \bar{a}(z, b, -\lambda)$. Then $\mathcal{C}^* W_{ir_j} = W_{ir_j}$.

Wigner distributions are fundamental in the theory of quantum ergodicity. Let us recall the basic result:

Theorem 3.7. [Sh, Z] *Let $d\mu$ denote Haar measure on $S\mathbf{X}_\Gamma$. Then*

$$\frac{1}{N(\lambda)} \sum_{j:|r_j|\leq\lambda} |\langle a, W_{ir_j}\rangle_{S\mathbf{X}_\Gamma} - \frac{1}{\mu(S\mathbf{X}_\Gamma)}\langle a, \mu\rangle_{S\mathbf{X}_\Gamma}|^2 \to 0,$$

where $N(\lambda)$ is the normalization factor $\sharp\{j : |r_j| \leq \lambda\}$.

It follows that a subsequence (W_{j_k}) of density one of the Wigner distributions tends to Liouville measure (which equals normalized Haar measure in this case). The "quantum unique ergodicity" problem is to know whether there exist exceptional subsequences with other limits. E. Lindenstrauss proved that no such exceptional sequences exist in the case of Hecke eigenfunctions on arithmetic surfaces [L]. In constant curvature -1 but without any arithmeticity assumption, Anantharaman–Nonnenmacher [AN] prove that the entropy of any quantum limit must be greater that $\frac{1}{2}$; although the methods in [AN] are rather disjoint from ours, it is no coincidence that the quantity $\frac{1}{2}$ is the same as $\Re e(\frac{1}{2} + ir_j)$.

4. Proof of Theorem 1.1

4.1. The operator L_r

We begin the proof with a lemma giving the explicit expression of W_{ir_j}:

Lemma 4.1. *We have*

$$\begin{aligned}&\langle Op(a)\phi_{ir_j}, \phi_{ir_j}\rangle_{L^2(\mathbf{X}_\Gamma)}\\ &= 2^{(1+2ir_j)} \int_{B\times B} \left(\int_{\mathbf{D}} \chi a(z, b)[\cosh s_{b',b}(z)]^{-(1+2ir_j)} dVol(z)\right) \frac{T_{ir_j}(db) T_{ir_j}(db')}{|b - b'|^{1+2ir_j}},\end{aligned} \tag{4.1}$$

where $\cosh s_{b_1,b_2}(z)$ *is given by (4.2). The right hand side is independent of the choice of* χ.

Proof. By the generalized Poisson formula and the definition of $Op(a)$,

$$\langle Op(a)\phi_{ir_j}, \phi_{ir_j}\rangle = \int_{B\times B}\left(\int_{\mathbf{D}} \chi a(z,b) e^{(\frac{1}{2}+ir_j)\langle z,b\rangle} e^{(\frac{1}{2}+ir_j)\langle z,b'\rangle} dVol(z)\right) T_{ir_j}(db) T_{ir_j}(db').$$

We then use the following identity

Lemma 4.2. [N] *Let* $z \in D$, *let* $b_1, b_2 \in B$ *and let* $s_{b_1,b_2}(z)$ *denote the hyperbolic distance from* z *to the geodesic* γ_{b_1,b_2} *defined by* (b_1, b_2). *Then*

$$\cosh s_{b_1,b_2}(z) = \frac{2|z-b_1||z-b_2|}{|b_1-b_2|(1-|z|^2)}.$$

Combined with (3.10) and (3.8), we get

$$e^{\langle z,b\rangle} e^{\langle z,b'\rangle} = 4[\cosh s_{b',b}(z)]^{-2}|b-b'|^{-2}.$$

Raising both sides to the power $\frac{1}{2} + ir_j$ completes the proof. □

The next step is to analyze the integral operator

$$\int_{\mathbf{D}} \chi a(z,b) e^{(\frac{1}{2}+ir)\langle z,b\rangle} e^{(\frac{1}{2}+ir)\langle z,b'\rangle} dVol(z) = 2^{(1+2ir)} \int_{\mathbf{D}} \chi a(z,b)[\cosh s_{b',b}(z)]^{-(1+2ir)}|b-b'|^{-(1+2ir)} dVol(z). \quad (4.2)$$

In this paragraph – and later in the paper – we sometimes drop the j-indices of r_j, indexing the eigenfunctions by r instead.

If we drop the factor $2^{(1+2ir)}|b-b'|^{-(1+2ir)}$, the right side of 4.2 defines the operator $\mathcal{L}_r : C_c(\mathbf{D}) \to C(B \times B)$ by

$$\mathcal{L}_r(\chi a)(b', b) = \int_{\mathbf{D}} \chi a(z,b)[\cosh s_{b',b}(z)]^{-(1+2ir)} dVol(z). \quad (4.3)$$

We now rewrite the integral in terms of coordinates $z = (t, u)$ based on the geodesic $\gamma_{b',b}$, after which we can relate $\mathcal{L}_r$ with the operator in (1.6).

Given a geodesic $\gamma_{b',b}$, we work with special coordinates on $\mathbf{D}$ or $\mathbf{H}$, adapted to $\gamma_{b',b}$ as follows. We write $z = (t, u)$ where t measures arclength on $\gamma_{b',b}$ and u measures arclength on horocycles centered at b. More precisely, we denote by $g(b', b)$ the vector on $\gamma_{b',b}$ which is closest to the origin, and the coordinates (t, u) parametrizing z are defined by $(z, b) = g(b', b)a_t n_u$. For any given (b', b), the volume element of z is $dVol = dtdu$. The computation is most easily checked in the upper half plane, with $b = \infty, b' = 0$ and $g(b', b) = e = (i, \infty)$. Then $a_t n_u i = e^t(i + u)$. The area form is $\frac{dxdy}{y^2}$. Setting $y = e^t, x = ue^t$ shows that the area form equals $dtdu$.

We obtain

$$\mathcal{L}_r(\chi a)(b',b) = \int \cosh s_{b',b}(t,u)^{-(1+2ir)} \chi a(g(b',b)a_t n_u) du dt. \tag{4.4}$$

We further simplify as follows:

Lemma 4.3. *We have*

$$\mathcal{L}_r(\chi a)(b,b') = \int_{\mathbf{R}\times\mathbf{R}} (1+u^2)^{-(\frac{1}{2}+ir)} \chi a(g(b,b')a_t n_u) du dt.$$

Proof. We recall that $s_{b',b}(t,u)$ is the distance from the basepoint of $ga_t n_u$ to the geodesic generated by g in the hyperbolic plane $\mathbf{H} = G/K$. That distance depends only on u and has the value $\cosh s_{b',b}(t,u) = \sqrt{1+u^2}$.

□

Next, we further rewrite the operator $\mathcal{L}_r$ in terms of the operator L_r in (1.6):

Lemma 4.4. *We have:*

$$\langle Op(a)\phi_{ir}, \phi_{ir}\rangle_{L^2(\mathbf{X}_\Gamma)} = 2^{(1+2ir)} \int_G L_r(\chi a)(g) PS_{ir}(dg).$$

Proof. Lemma 4.3 states that

$$\begin{aligned}\mathcal{L}_r(\chi a)(b,b') &= \int_{\mathbf{R}} L_r(\chi a)(g(b,b')a_t) dt \\ &= \mathcal{R}(L_r(\chi a))(b,b').\end{aligned}$$

Integrating against dps_{ir} and using the formula in Definition 3.3 completes the proof.

□

The next step is to apply the stationary phase method to $L_r(\chi a)$. The stationary phase set of (4.4) is the geodesic $\gamma_{b',b}$ from b' to b or equivalently it is the set $u = 0$ in the integral defining $L_r(\chi a)$. Since $\big(\log(1+u^2)''\big)\,|_{u=0} = 2$, the stationary phase method gives the asymptotic expansion

$$L_r(\chi a)(g) = (-ir/\pi)^{-1/2}\big(\sum_{n\geq 0} r^{-n} L_{2n}(\chi a)(g)\big) \tag{4.5}$$

where L_{2n} is a differential operator of order $2n$ on $S\mathbf{D}$: L_0 is the identity, the other L_{2n} are differential operators in the stable direction, that is, in the direction n_u generated by the vector field X_+.

If we now integrate (4.5) with respect to PS_{ir}, and compare with Lemma 4.4, we get an asymptotic expansion,

$$\langle Op(a)\phi_{ir}, \phi_{ir}\rangle_{L^2(\mathbf{X}_\Gamma)} = 2^{(1+2ir)}(-ir/\pi)^{-1/2}\big(\sum_{n\geq 0} r^{-n} \int_{S\mathbf{D}} L_{2n}(\chi a)(g) PS_{ir}(dg)\big) \tag{4.6}$$

Because the distribution on the left-hand side, $e^{(\frac{1}{2}+ir)\langle z,b\rangle}e^{(\frac{1}{2}+ir)\langle z,b'\rangle}dVol(z)$ $T_{ir}(db)T_{ir}(db')$, is Γ-invariant (as a distribution in the triple (b,b',z)), each of the distributions obtained in the expansion,

$$f \mapsto \int_{S\mathbf{D}} L_{2n}(f)(g)PS_{ir}(dg),$$

is Γ-invariant. In application of the principle 3.5, the functional

$$a \mapsto \int_G L_{2n}(\chi a)(g)PS_{ir}(dg)$$

defines a distribution on $\Gamma\backslash G$, and the definition does not depend on the choice of χ. The first term ($n=0$) is precisely the Patterson-Sullivan distribution PS_{ir} as defined in the quotient $S\mathbf{X}_\Gamma$.

4.2. Completion of Proof of Theorem 1.1

We now turn to the relation between W_{ir} and PS_{ir}. It follows from the stationary phase asymptotics above, (4.5), that

$$\int_{S\mathbf{X}_\Gamma} a(g)W_{ir_j}(dg) = \\ 2^{(1+2ir)}(-ir/\pi)^{-1/2}\sum_{n=0}^{N} r_j^{-n}\int_{S\mathbf{D}} L_{2n}(\chi a)(g)PS_{ir_j}(dg) + O(r_j^{-N-1+K})$$

where K was defined in 3.14. If we choose $N>K$ then the remainder term goes to zero. Since $L_0 = Id$, the operator $L_r^{(N)} = \sum_{n=0}^N r^{-n}L_{2n}$ can be inverted up to $O(r^{-N-1})$, that is, one can find differential operators $M_r^{(N)} = \sum_{n=0}^N r^{-n}M_{2n}$ (with $M_0 = Id$) and $R_r^{(N)}$ such that

$$L_r^{(N)}M_r^{(N)} = Id + r^{-N-1}R_r^{(N)}.$$

We thus get

$$\begin{aligned}\int_{S\mathbf{X}_\Gamma} M_{r_j}^{(N)}a(g)W_{ir_j}(dg) &= \int_{S\mathbf{D}} L_{r_j}^{(N)}\chi M_{r_j}^{(N)}a(g)PS_{ir_j}(dg) + O(r_j^{-N-1+K}) \\ &= \int_{S\mathbf{D}} L_{r_j}^{(N)}M_{r_j}^{(N)}\chi a(g)PS_{ir_j}(dg) + O(r_j^{-N-1+K}) \\ &= \int_{S\mathbf{X}_\Gamma} a(g)PS_{ir_j}(dg) + O(r_j^{-N-1+K})\end{aligned}$$

The second line is a consequence of Lemma 3.5. Since we know, from standard estimates on pseudo-differential operators, that the Wigner measures are uniformly bounded in $(C^k)^*$ for some k, we have

$$\int_{S\mathbf{X}_\Gamma} M_{r_j}^{(N)}a(g)W_{ir_j}(dg) = \int_{S\mathbf{X}_\Gamma} a(g)W_{ir_j}(dg) + O(r_j^{-1}).$$

This shows that

$$2^{(1+2ir_j)}(-ir_j/\pi)^{-1/2}\int_{S\mathbf{X}_\Gamma} a(g)PS_{ir_j}(dg) = \int_{S\mathbf{X}_\Gamma} a(g)W_{ir_j}(dg) + O(r_j^{-1}).$$

The left side must be asymptotically the same as $\langle a, \widehat{PS}_{ir_j}\rangle$ since the leading coefficients must match when $a = 1$. This completes the proof of Theorem 1.1.

Remark 4.1. One can directly show that the coefficient on the left side is asymptotically the same as the normalizing factor $2^{(1+2ir_j)}\mu_0(\frac{1}{2}+ir_j)$ by using properties of the Γ function. It suffices to show

$$2^{1+2ir_j}(-ir_j/\pi)^{-1/2} \sim 2^{(1+2ir_j)}\mu_0(\frac{1}{2}+ir_j),$$

which follows from the standard fact that

$$\frac{\Gamma(\frac{1}{2})\Gamma(ir_j)}{\Gamma(\frac{1}{2}+ir_j)} \sim \pi^{1/2}(-ir_j)^{-1/2}.$$

The agreement is not surprising, since the last evaluation can be obtained by applying the stationary phase method as in the proof of Theorem 1.1 to the integral $\int_{\mathbf{R}}(1+u^2)^{-(\frac{1}{2}+ir)}du$.

5. Integral operators and eigenfunctions

In this section, we give further results on the operators L_r (1.6) and $I_{PS_{ir}}$ (1.8) which will be needed in the proof of Theorem 1.3. With no extra work, we treat general integral operators of the form

$$I_\mu(\sigma)(u) := \int_{\Gamma\backslash G} \sigma(gn_u)\mu(dg), \tag{5.1}$$

where $\sigma \in C^\infty(\Gamma\backslash G)$ is an automorphic form and where μ is an invariant distribution for the geodesic flow on $\Gamma\backslash G$. In addition to $\mu = PS_{ir}$ the case where μ is a periodic orbit measure is also important in this article. In this case, we write $I_\mu = I_\gamma$ with $I_\gamma(\sigma)(u) = \int_{\langle L_\gamma\rangle\backslash A}\sigma(\alpha_\gamma^{-1}an_u)da$. Here, $\alpha_\gamma \in G$ is an element conjugating $\gamma \in \Gamma$ to an element of A. This expression arose in the trace formulae of [Z] and will arise in §9. The similarity of these two kinds of integral operators may be seen as one of the deus ex machina behind Theorem 1.3.

5.1. The integral operator I_μ

We can view I_μ as an integral operator from $C^\infty(\Gamma\backslash G) \to C^\infty(N) \simeq C^\infty(\mathbf{R})$. The following lemma shows that when σ is a joint eigenfunction of the Casimir operator and of W, then $I_\mu(\sigma)$ solves an ordinary differential equation in u. When σ is a (τ, m)-eigenfunction in the complementary or principal unitary series, the equation is

$$(u^2+1)\frac{d^2f}{du^2} + (2u - im)\frac{df}{du} + (\frac{1}{4}+\tau^2)f = 0 \tag{5.2}$$

We denote by $F_{\tau,m}\left(\frac{u-i}{-2i}\right)$ the even solution of (5.2) which equals 1 at $u=0$, and by $G_{\tau,m}\left(\frac{u-i}{-2i}\right)$ the odd solution whose derivative equals 1 at $u=0$. In the holomorphic discrete series, and when σ is the lowest weight vector, the analogous equation is the first order equation

$$2i\frac{df}{du} = (-2\frac{d}{du} - m)f. \tag{5.3}$$

A basis for its solutions is given by $f(u) = (-i)^{-m/2}(u+i)^{-m/2}$. There are similar equations for higher weights and for the anti-holomorphic discrete series, but for simplicity we only discuss the lowest weight case.

Proposition 5.1. *Let μ be a geodesic flow invariant distribution on $\Gamma\backslash G$.*

- *Let σ be a (τ, m)-eigenfunction in the principal or complementary series. Then $I_\mu(\sigma)(u)$ is a solution of (5.2). Hence,*

$$I_\mu(\sigma)(u) = \langle\sigma,\mu\rangle_{\Gamma\backslash G}\, F_{\tau,m}\left(\frac{u-i}{-2i}\right) + \langle X_+\sigma,\mu\rangle_{\Gamma\backslash G}\, G_{\tau,m}\left(\frac{u-i}{-2i}\right),$$

  where F, G are the fundamental solutions of (5.2) defined in [Z] *(2.3) (see (5.7) for formulae in terms of hypergeometric functions).*
- *Let σ be a (τ, m)-eigenfunction in the discrete holomorphic or anti-holomorphic series. For simplicity, assume $\sigma = \psi_m$ (the lowest weight vector in $\mathcal{D}_m^+$). Then:*

$$I_\mu(\sigma)(u) = \langle\sigma,\mu\rangle_{\Gamma\backslash G}\,(-i)^{-m/2}(u+i)^{-m/2}.$$

Proof. In the case of $I_\mu = I_\gamma$, the proof is given in [Z], Proposition 2.3 and Corollary 2.4. We briefly verify that the same proof works for any invariant distribution.

First assume σ is a (τ, m)-eigenfunction in the continuous series. Since $4\Omega = H^2 + 4X_+^2 - 2H - 4X_+W$ we find that

$$\left(4\frac{d^2}{du^2} - 4im\frac{d}{du} + 4(\frac{1}{4}+\tau^2)\right) I_\mu(\sigma)(u) = -\int_{\Gamma\backslash G}\big((H^2-2H)\sigma\big)(gn_u)\mu(dg). \tag{5.4}$$

We write $H\sigma(g)$ as $2\frac{d}{dt}_{t=0}\sigma(ga_t)$. Using that $n_u a_t = a_t n_{ue^{-t}}$ and that μ is an A-invariant distribution, we find that $\int H\sigma(ga_t)\mu(dg) = -2u\frac{d}{du}I_\mu(\sigma)(u)$. A similar calculation replaces H^2 by the square of this operator. The final equation is as stated above. We then evaluate $I_\mu(\sigma)$ and its first derivative at $u=0$ to obtain the expression in terms of F, G.

In the discrete holomorphic series, we use that $E_-\sigma = 0$ to get $2iX_+\sigma = (H-m)\sigma$. This leads to equation (5.3) and to the solution given above.

□

5.2. The integral $\int_{\mathbf{R}}(1+u^2)^{-s}I_\mu(\sigma)(u)du$

In Theorems 1.2, 1.3 and elsewhere, we will need explicit formulae for the integrals

$$\int_{\mathbf{R}}(1+u^2)^{-s}I_\mu(\sigma)(u)du \tag{5.5}$$

We assemble the results here for future reference.

In view of Proposition 5.1, we need explicit formulae for the integral of $(1+u^2)^{-s}$ against the functions $F_{\tau,m}\left(\frac{u-i}{-2i}\right), G_{\tau,m}\left(\frac{u-i}{-2i}\right)$, and $(-i)^{-m/2}(u+i)^{-m/2}$. In fact, by Proposition 2.2 and Lemma 2.3, it will suffice for Theorems 1.2 and 1.3 to have explicit formulae just for $F_{\tau,0}$ and $(-i)^{-m/2}(u+i)^{-m/2}$.

We introduce the following notation:

$$\begin{cases} \mu_0(s) & := \int_0^\infty (u^2+1)^{-s}du, \\ \mu^c_{ir_k}(s) & := \int_{\mathbf{R}}(u^2+1)^{-s}F(\frac{1}{4}+\frac{2ir_k}{4},\frac{1}{4}-\frac{2ir_k}{4},\frac{1}{2},-u^2)du, \\ \mu^c_{\tau,m}(s) & := \int_{-\infty}^\infty (u^2+1)^{-s}F_{\tau,m}(\frac{u-i}{-2i})du, \\ \mu^{codd}_{\tau,m}(s) & := \int_{\mathbf{R}}(u^2+1)^{-s}G_{\tau,m}\left(\frac{u-i}{-2i}\right)du, \\ \mu^d_m(s) & := \int_{\mathbf{R}}(u+i)^{-m/2}(u^2+1)^s du, \ ; \end{cases} \tag{5.6}$$

It is clear that the integrals defining $\mu_0(s)$ and $\mu^d_m(s)$ converge absolutely for $\Re e\ s>\frac{1}{2}$ and $\Re e(2s-\frac{m}{2})<-1$, respectively. We now show:

Proposition 5.2. *The integral defining $\mu^c_{ir_k}(s)$ converges absolutely for $-2\Re es-1/2+\Re e(ir_k)<-1$, and in this region we have:*

$$|\mu^c_{ir_k}(s)|\le C\int_{-\infty}^\infty (|u|+1)^{-2\Re es-1/2+\Re e(ir_k)}\,du,$$

for some constant C (independent of s, r_k).

Proof. Indeed, as in [Z] (Proposition 2.7), the differential equation (5.2) is equivalent, by a change of variables, to a hypergeometric equation, and a short calculation shows that

$$\begin{cases} F_{ir_k,0}\left(\frac{u-i}{-2i}\right) = F(\frac{1}{4}+\frac{2ir_k}{4},\frac{1}{4}-\frac{2ir_k}{4},\frac{1}{2},-u^2), \\ G_{ir_k,0}\left(\frac{u-i}{-2i}\right) = (-2iu)F(\frac{3}{4}+\frac{2ir_k}{4},\frac{3}{4}-\frac{2ir_k}{4},\frac{3}{2},-u^2). \end{cases} . \tag{5.7}$$

Classical estimates on hypergeometric functions (see also [Z], p. 50) show that there exists $C>0$ (independent of r_k) such that

$$\begin{cases} \left|F(\frac{1}{4}+\frac{2ir_k}{4},\frac{1}{4}-\frac{2ir_k}{4},\frac{1}{2},-u^2)\right| \\ \left|uF(\frac{3}{4}+\frac{2ir_k}{4},\frac{3}{4}-\frac{2ir_k}{4},\frac{3}{2},-u^2)\right| \end{cases} \le C\ (1+|u|)^{-1/2+\Re e ir_k}, \tag{5.8}$$

These estimates follow immediately from the connection formulae for hypergeometric functions:

$$\begin{aligned} F(a,b,c,z) &= \tfrac{\Gamma(c)\Gamma(b-a)}{\Gamma(b)\Gamma(c-a)}(-z)^{-a}F(a,1+a-c,1+a-b;z^{-1}) \\ &\quad + \tfrac{\Gamma(c)\Gamma(a-b)}{\Gamma(a)\Gamma(c-b)}(-z)^{-b}F(b,1+b-c,1+b-a;z^{-1}). \end{aligned}$$

Since $F(0)=1$, we obtain that (as $|u|\to\infty$)

$$\begin{aligned} F(\tfrac14+\tfrac{2ir_k}{4},\tfrac14-\tfrac{2ir_k}{4},\tfrac12,-u^2) &\sim \tfrac{\Gamma(\frac12)\Gamma(-ir_k)}{\Gamma(\frac14-\frac{ir_k}{2})^2}|u|^{-(\frac12+ir_k)} \\ &\quad + \tfrac{\Gamma(\frac12)\Gamma(ir_k)}{\Gamma(\frac14+\frac{2ir_k}{4})^2}|u|^{-(\frac12-ir_k)}. \end{aligned} \tag{5.9}$$

The asymptotics (2.10) of the Γ function on vertical lines shows that the ratios of Γ functions are uniformly bounded in r_k. The decay rate $|u|^{-(\frac12-ir_k)}$ is sufficient for the absolute convergence of the integral in (5.6) as long as $\Re(\frac12-ir_k)>0$, i.e. if ir_k is not the parameter of the trivial representation.

□

Although we will not need them, we note that the estimates for G are similar. Each of the above functions admits a meromorphic continuation to **C**. Since we will not need the results for general $\mu^c_{\tau,m}(s), \mu^{codd}_{\tau,m}(s)$ we omit them in the following.

Proposition 5.3. *We have:*

$$\begin{cases} \mu_0(s) = \frac{\Gamma(\frac12)\Gamma(s-\frac12)}{\Gamma(s)} & (\Re e\ s>\frac12) \\[2mm] \mu^c_{ir_k}(s) = \frac{\Gamma(\frac12)\Gamma(s-\frac14+\frac{2ir_k}{4})\Gamma(s-\frac14-\frac{2ir_k}{4})}{\Gamma(s)^2} & (\Re e\ s>0) \\[2mm] \mu^d_m(s) = \frac{(-i)^{m/2}\pi 2^{2s+2-m/2}\Gamma(-2s+\frac m2)}{-(2s+1-\frac m2)\Gamma(-s)\Gamma(-s+\frac m2)} & (\Re(2s-\frac m2)<-1). \end{cases}$$

The proof is given in [Z] (see pages 50-52).

6. Proof of Theorem 1.2.

The key objects in the proof of Theorem 1.2 are the closely related integrals

$$\begin{cases} I_r(\sigma) = \int_{\mathbf{R}}(1+u^2)^{-(\frac12+ir)}\langle(\sigma\chi)^u, PS_{ir}\rangle_{S\mathbf{D}}\,du, \\[2mm] I^\Gamma_r(\sigma) = \int_{\mathbf{R}}(1+u^2)^{-(\frac12+ir)}\langle\sigma^u, PS_{ir}\rangle_{S\mathbf{X}_\Gamma}\,du \end{cases} \tag{6.1}$$

where PS_{ir_j} is defined in Definition 3.3 as a distribution on $S\mathbf{D}$ or on the quotient $S\mathbf{X}_\Gamma$, and where $f^u(g)=f(gn_u)$. Note that $\langle\sigma^u, PS_{ir}\rangle_{S\mathbf{X}_\Gamma} = I_{PS_{ir}}(\sigma)(u)$ in the notation of §5. It takes some work to prove that each integral is well-defined. In Lemma 6.1 it is proved that the two integrals are well-defined and equal for $\sigma\in C^\infty(\Gamma\backslash G)$.

Theorem 1.1 equates the Wigner distribution with the distribution $\sigma \to \langle L_{r_j}(\chi\sigma), PS_{ir_j}\rangle_{S\mathbf{D}}$ for $\sigma \in C^\infty(\Gamma\backslash G)$. In Proposition 6.4 we show that this functional also equals $I_r(\sigma) = I_r^\Gamma(\sigma)$. The explicit formulae for the Wigner distributions in terms of the Patterson-Sullivan distributions follow from the identification with $I_r^\Gamma(\sigma)$, which can be explicitly evaluated using the results of §5.

6.1. Convergence and equality of the integrals

In the following, we recall that $\Re e(ir_k) = 0$ in the unitary principal series but is positive in the complementary series.

Proposition 6.1. *We have:*

1. *If $\mathcal{P}_{ir}$ is in the unitary principal series and $\sigma \in C^\infty(\Gamma\backslash G)$ is orthogonal to constant functions, then the integral $I_r^\Gamma(\sigma)$ converges absolutely.*
2. *Under the same assumptions we have $I_r(\sigma) = I_r^\Gamma(\sigma)$.*

6.1.1. Proof of (1). We give a representation theoretic proof that

$$\int_{\mathbf{R}} (1+u^2)^{-(\frac{1}{2}+ir)} I^\Gamma_{PS_{ir}}(\sigma)(u)du \tag{6.2}$$

converges absolutely. We make no attempt at a sharp estimate but only one sufficient for the purposes of this paper.

Lemma 6.2. *Let PS_{ir} be the Patterson-Sullivan distribution corresponding to ϕ_r. Then:*

$$\begin{cases} (i) \quad I_{PS_{ir_j}}(\phi_{ir_k})(u) \leq C\,(1+|r_k|)^4(1+|r_j|)^K(1+|u|)^{-1/2+\Re(ir_j)}; \\ \\ (ii) \quad I_{PS_{ir_j}}(\psi_m) \leq C\,(1+|m|)^4(1+|r_j|)^K(1+|u|)^{-m/2}, \end{cases}$$

*where K is the same as in (*3.14*).*

Proof. (i) By Propositions 5.1 and 2.3,

$$\begin{aligned} I_{PS_{ir_j}}(\phi_{ir_k})(u) &= \textstyle\int_{\Gamma\backslash G} \phi_{ir_k}(gn_u)PS_{ir_j}(dg) \\ &= \big(\textstyle\int_{\Gamma\backslash G} \phi_{ir_k}PS_{ir_j}(dg)\big)F_{ir_k,0}\left(\frac{u-i}{-2i}\right) \end{aligned} \tag{6.3}$$

By (3.14), there exists K so that

$$|\langle \phi_{ir_k}, PS_{ir_j}\rangle| \leq C(1+|r_j|)^K(1+|r_k|)^4.$$

Here, we used a crude estimate $||\phi_{ir_k}||_{C^2} \leq C(1+|r_k|)^4$ (in fact, $r_k^3/\log r_k$ is true, but it is not necessary for our argument). We combine with the estimates in Proposition 5.2 (cf. 5.9) on the hypergeometric factor to obtain the estimate stated in (i).

(ii) We now have

$$\begin{aligned} I_{PS_{ir_j}}(\psi_m)(u) &= \textstyle\int_{\Gamma\backslash G} \psi_m(gn_u)PS_{ir_j}(dg) \\ &= \big(\textstyle\int_{\Gamma\backslash G} \psi_m PS_{ir_j}(dg)\big)(u+i)^{-m/2}. \end{aligned} \tag{6.4}$$

To complete the proof we note that $|(u+i)^{-m/2}| \leq C(1+|u|)^{-m/2}$ and that (by (3.14)),

$$|\langle \psi_m, PS_{ir_j} \rangle| \leq C(1+|r_j|)^K (1+|m|)^4.$$

□

Given a co-compact discrete group $\Gamma \subset SL(2,\mathbf{R})$ we denote by $\tau_0 = \Re e(ir_0)$ the real part of the Casimir parameter corresponding to the lowest non-zero eigenvalue of $\triangle$, i.e. the complementary series representation closest to the trivial representation.

Lemma 6.3. *If $\sigma \in C^\infty(\Gamma\backslash G)$ has no component in the trivial representation, we have:*

$$I_{PS_{ir}}(\sigma)(u) \leq C(1+|r|)^K (1+|u|)^{-1/2+\tau_0}.$$

Proof. Since PS_{ir} is geodesic flow and time reversal invariant, we may write by Proposition 2.4,

$$I_{PS_{ir}}(\sigma)(u) = \sum_{r_j} \frac{\langle \sigma, \Xi_{ir_j} \rangle}{\langle \phi_{ir_j}, \Xi_{ir_j} \rangle} I_{PS_{ir}}(\phi_{ir_j})(u) + \sum_{m,\pm} \frac{\langle \sigma, \Xi_m^\pm \rangle}{\langle \psi_m, \Xi_m^\pm \rangle} I_{PS_{ir}}(\psi_m)(u). \quad (6.5)$$

It follows by Lemma 6.2 that

$$\begin{aligned} |I_{PS_{ir}}(\sigma)(u)| \quad &\leq \quad C(1+|r|)^K \times \{\sum_{r_j} (1+|r_j|)^4 \left| \frac{\langle \sigma, \Xi_{ir_j} \rangle}{\langle \phi_{ir_j}, \Xi_{ir_j} \rangle} \right| (1+|u|)^{-1/2+\Re e(ir)} \\ &+ \quad \sum_m (1+|m|)^4 |\frac{\langle \sigma, \Xi_m^\pm \rangle}{\langle \psi_m, \Xi_m^\pm \rangle}| \ (1+|u|)^{-m/2}\} \ . \end{aligned} \quad (6.6)$$

By Proposition 2.5,

$$\left| \frac{\langle \sigma, \Xi_{ir_j} \rangle}{\langle \phi_{ir_j}, \Xi_{ir_j} \rangle} \right| \leq ||X_- \mathcal{P}_{ir_j}(\sigma)||_{L^2}, \quad |\frac{\langle \sigma, \Xi_m^\pm \rangle}{\langle \psi_m, \Xi_m^\pm \rangle}| \leq ||X_- \mathcal{D}_m^\pm(\sigma)||_{L^2}.$$

It follows that for any $M > 0$ there exists a constant C_M so that

$$\left| \frac{\langle \sigma, \Xi_{ir_j} \rangle}{\langle \phi_{ir_j}, \Xi_{ir_j} \rangle} \right| \leq C_M (1+|r_j|)^{-M}, \quad |\frac{\langle \sigma, \Xi_m^\pm \rangle}{\langle \psi_m, \Xi_m^\pm \rangle}| \leq C_M (1+|m|)^{-M}. \quad (6.7)$$

Indeed,

$$\mathcal{P}_{ir}(\sigma) = \sum_{m \in \mathbf{Z}} \sigma_{ir,m} \phi_{ir,m}, \quad \text{with} \quad |\sigma_{ir,m}| \leq C_M (1+|r_j|+|m|)^{-M},$$

hence

$$||X_- \mathcal{P}_{ir_j}(\sigma)||_{L^2} \leq C_M \sum_m (1+|r_j|+|m|)^{-M} (1+|r_j|+|m|),$$

where we bound $||X_- \phi_{ir,m}||_{L^2} \leq C(1+|r_j|+|m|)$.

Similarly,

$$(\mathcal{D}_m^\pm)(\sigma) = \sum_{n=0}^\infty \sigma_{m,m+2n} \psi_{m,m+2n}, \quad \text{with} \quad |\sigma_{m,m+2n}| \leq C_M (1+|m|+|n|)^{-M},$$

hence

$$||X_-\mathcal{D}^{\pm}_m(\sigma)||_{L^2} \leq C_M \sum_m (1+|m|+|n|)^{-M}(1+|m|+|n|).$$

By (6.7) and Lemma (6.2), the sum (6.5) converges absolutely and the decay estimates in u sum up to the stated rate. □

Completion of proof of Proposition 6.1 (1): It follows from Lemma 6.3 that

$$\begin{aligned} |I^{\Gamma}_r(\sigma)| &\leq \int_{\mathbf{R}} (1+u^2)^{-(\frac{1}{2}+\Re e ir)} |I_{PS_{ir}}(\sigma)(u)| du \\ &\leq C(1+|r|)^K \int_{\mathbf{R}} |(1+u^2)^{-(\frac{1}{2}+\Re e(ir))}| (1+|u|)^{-1/2+\tau_0} du. \end{aligned} \tag{6.8}$$

Since $\mathcal{P}_{ir}$ is in the unitary principal series, $\Re e(ir) = 0$ and so $|(1+u^2)^{-(\frac{1}{2}+ir)}| = (1+u^2)^{-\frac{1}{2}}$ and since $-1/2+\tau_0 < 0$ it follows that the last integral in (6.8) converges absolutely.

We now move on to the assertion (2) of Proposition 6.1.

6.1.2. Proof of (2). By Proposition 3.5, we have

$$\int_G \sigma(gn_u)\chi(gn_u) PS_{ir}(dg) = \int_G \sigma(gn_u)\chi(g) PS_{ir}(dg). \tag{6.9}$$

Indeed, $\chi(g)$ and $\chi^u(g) := \chi(gn_u)$ are both smooth fundamental cutoffs, so both sides equal $\langle \sigma, PS_{ir}\rangle_{\Gamma\backslash G}$. Integrating against $\int_{\mathbf{R}}(1+u^2)^{-(\frac{1}{2}+ir)}$ completes the proof.

6.2. Continuity of PS_{ir}

As mentioned above, the Wigner distribution equals the functional $\sigma \to \langle L_{r_j}(\chi\sigma), PS_{ir_j}\rangle_{S\mathbf{D}}$. To prove that this also equals $I_r(\sigma) = I^{\Gamma}_r(\sigma)$ we need the following continuity result for the functional PS_{ir}.

Lemma 6.4. *$PS_{ir} \in \mathcal{D}'(S\mathbf{D})$ has the following continuity property,*

$$\langle L_{r_j}(\chi\sigma), PS_{ir_j}\rangle_{S\mathbf{D}} = \int_{\mathbf{R}} (1+u^2)^{-(\frac{1}{2}+ir)} \langle (\sigma\chi)^u, PS_{ir}\rangle_{S\mathbf{D}}\, du,$$

where $f^u(g) = f(gn_u)$.

Proof. By Definition 3.3,

$$\begin{aligned} \langle (\sigma\chi)^u, PS_{ir}\rangle_{S\mathbf{D}} &= \langle \mathcal{R}(\sigma\chi)^u, ps_{ir}\rangle_{B\times B} \\ &= \int_{B\times B'} \left\{ \int_{\mathbf{R}_t} (\chi\sigma)(g(b,b')a_t n_u) dt \right\} \frac{T_{ir}(db)T_{ir}(db')}{|b-b'|^{1+2ir}} \end{aligned}$$

We first note that for all u, $\mathcal{R}(\chi\sigma)^u \in C^\infty_c(B\times B\backslash\Delta)$ since $(\chi\sigma)^u \in C^\infty_c(S\mathbf{D})$. It follows that $ps_{ir}(\mathcal{R}(\sigma\chi)^u)$ is well-defined and smooth in u.

The continuity statement is equivalent to

$$\langle \mathcal{R}L_r(\chi\sigma), ps_{ir}\rangle_{B\times B} = \int_{\mathbf{R}} (1+u^2)^{-(\frac{1}{2}+ir)} \langle \mathcal{R}(\sigma\chi)^u(b,b'), ps_{ir}\rangle_{B\times B}\, du, \tag{6.10}$$

or equivalently

$$\left\langle \int_{\mathbf{R}} (1+u^2)^{-(\frac{1}{2}+ir)} \mathcal{R}(\sigma\chi)^u(b,b')du, ps_{ir} \right\rangle_{B\times B} = \int_{\mathbf{R}} (1+u^2)^{-(\frac{1}{2}+ir)} \langle \mathcal{R}(\sigma\chi)^u(b,b'), ps_{ir}\rangle_{B\times B}\, du. \quad (6.11)$$

We must again check that both sides of (6.10) are well-defined. Clearly, $\mathcal{R}(L_r(\chi\sigma))$ is well-defined because χ has compact support. The problem is to prove that the left-hand side is well-defined, since that ps_{ir} is only known to be a bounded linear functional on $|b-b'|^{1+2ir}C^2(B\times B)$ (cf. Proposition 3.6). We therefore have to verify that

$$\int_{\mathbf{R}} (1+u^2)^{-(\frac{1}{2}+ir)} \mathcal{R}(\sigma\chi)^u(b,b')du \in |b-b'|^{1+2ir}C^2(B\times B).$$

By Lemma 4.3 and (4.3), we have

$$\int_{\mathbf{R}} (1+u^2)^{-(\frac{1}{2}+ir)} \mathcal{R}(\sigma\chi)^u(b,b')du = |b-b'|^{(1+2ir)} \int_D (\chi\sigma)(z,b) e^{(\frac{1}{2}+ir)\langle z,b\rangle} e^{(\frac{1}{2}+ir)\langle z,b'\rangle} dVol(z), \quad (6.12)$$

and therefore the condition to be satisfied is that

$$\int_D (\chi\sigma)(z,b) e^{(\frac{1}{2}+ir)\langle z,b\rangle} e^{(\frac{1}{2}+ir)\langle z,b'\rangle} dVol(z) \in C^\infty(B\times B). \quad (6.13)$$

This is clear due to the compact support of $\chi\sigma$ in z, which is independent of (b,b').

We may then rewrite (6.11) as:

$$\left\langle \Big(\int_D (\chi\sigma)(z,b) e^{(\frac{1}{2}+ir)\langle z,b\rangle} e^{(\frac{1}{2}+ir)\langle z,b'\rangle} dVol(z)\Big), T_{ir}\otimes T_{ir} \right\rangle_{B\times B} = \int_D \chi \langle \big(\sigma(z,b) e^{(\frac{1}{2}+ir)\langle z,b\rangle} e^{(\frac{1}{2}+ir)\langle z,b'\rangle}\big), T_{ir}\otimes T_{ir}\rangle_{B\times B}\, dVol(z). \quad (6.14)$$

I.e. we need to check that we can pass $T_{ir}\otimes T_{ir}$ under the $dVol(z)$ integral sign.

By Otal's regularity theorem (see Theorem 3.2), $T_{ir}(db) = F'_{ir}(b)db$ where F_{ir} is a continuous 2π periodic function in the sense that $F_{ir}(\theta+2\pi) - F_{ir}(\theta) = C_r$. Integration by parts then gives

$$\langle g, T_{ir}\rangle_B = \int_B g(b)T_{ir}(db) = -\int_B g'(b)F_{ir}(b)db + g(0)(F_{ir}(2\pi) - F_{ir}(0)).$$

Applying this in each of the (b,b') variables to the pairing on $B\times B$ in (6.14) produces four terms of which three involve the boundary term $(F_{ir}(2\pi) - F_{ir}(0))$ and the fourth is

$$\int_{B\times B} \{ \int_D \frac{\partial^2}{\partial b \times \partial b'} \Big((\chi\sigma)(z,b) e^{(\frac{1}{2}+ir)\langle z,b\rangle} e^{(\frac{1}{2}+ir)\langle z,b'\rangle} dVol(z) \Big) \} F_{ir}(b)F_{ir}(b')dbdb'.$$

By applying Fubini's theorem to the fourth term, we obtain:

$$\int_{B\times B}\{\int_D \frac{\partial^2}{\partial b\times\partial b'}\Big((\chi\sigma)(z,b)e^{(\frac{1}{2}+ir)\langle z,b\rangle}e^{(\frac{1}{2}+ir)\langle z,b'\rangle}dVol(z)\Big)\}F_{ir}(b)F_{ir}(b')dbdb'$$

$$= \int_D \chi\{\int_{B\times B}\frac{\partial^2}{\partial b\times\partial b'}\Big(\sigma(z,b)e^{(\frac{1}{2}+ir)\langle z,b\rangle}e^{(\frac{1}{2}+ir)\langle z,b'\rangle}\Big)F_{ir}(b)_{ir}(b')dbdb'\}dVol(z). \tag{6.15}$$

Fubini's theorem applies in a similar way to the other terms. We then transfer the b derivatives back to T_{ir} and obtain (6.14).

□

As a corollary of Proposition 6.1, we obtain the following explicit formula:

Corollary 6.5. *We have:*

$$\begin{aligned}\int_{\mathbf{R}}(1+u^2)^{-(\frac{1}{2}+ir)}I^{\Gamma}_{PS_{ir}}(\sigma)(u)du \quad &= \\ &\sum_{r_j}\frac{\langle\sigma,\Xi_{ir_j}\rangle}{\langle\phi_{ir_j},\Xi_{ir_j}\rangle}\Big(\int_{\Gamma\backslash G}\phi_{ir_j}PS_{ir}(dg)\Big)\mu^c_{ir_j}(\frac{1}{2}+ir)\\ &+\sum_{m,\pm}\frac{\langle\sigma,\Xi^{\pm}_m\rangle}{\langle\psi_m,\Xi^{\pm}_m\rangle}\Big(\int_{\Gamma\backslash G}\psi_m PS_{ir}(dg)\Big)\mu^d_m(\frac{1}{2}+ir).\end{aligned} \tag{6.16}$$

All integrals and series converge absolutely.

Proof. In fact, by Lemma 6.3 we may interchange the order of summation in (6.5) and integration in (6.2). Using (6.3) and (5.6), we have

$$\int_{\mathbf{R}}(1+u^2)^{-(\frac{1}{2}+ir)}I^{\Gamma}_{PS_{ir}}(\phi_{ir_j})(u)du = \Big(\int_{\Gamma\backslash G}\phi_{ir_j}PS_{ir}(dg)\Big)\mu^c_{ir_j}(\frac{1}{2}+ir),$$

and thus obtain the first series of the stated formula. Using (6.4) and (5.6), we have

$$\int_{\mathbf{R}}(1+u^2)^{-(\frac{1}{2}+ir)}I^{\Gamma}_{PS_{ir}}(\psi_m)(u)du = \Big(\int_{\Gamma\backslash G}\psi_m PS_{ir}(dg)\Big)\mu^d_m(\frac{1}{2}+ir),$$

and obtain the second series. □

6.3. Completion of Proof of Theorem 1.2

To complete the proof it suffices to explicitly evaluate $I^{\Gamma}_r(\sigma)$ for the generating automorphic forms.

Lemma 6.6. *In the special cases when $\sigma = \phi_{ir_k}, X_+\phi_{ir_k}$ or ψ_m, we have the explicit formulae:*

1. *In the case $\sigma = \phi_{ir_k}$, $\langle Op(\phi_{ir_k})\phi_{ir_j},\phi_{ir_j}\rangle = \mu^c_{ir_k}(\frac{1}{2}+ir_j)\left\langle\phi_{ir_k},PS_{ir_j}\right\rangle_{S\mathbf{X}_\Gamma}$.*
2. *For $\sigma = X_+\phi_{ir_k}$, $\langle Op(X_+\phi_{ir_k})\phi_{ir_j},\phi_{ir_j}\rangle = 0$ for all j.*
3. *For $\sigma = \psi_m$, $\langle Op(\psi_m)\phi_{ir_j},\phi_{ir_j}\rangle = \mu^d_m(\frac{1}{2}+ir_j)\left\langle\psi_m,PS_{ir_j}\right\rangle_{S\mathbf{X}_\Gamma}$.*

   Here, the expressions $\mu^c_{ir_k}(\frac{1}{2}+ir_j),\mu^d_m(\frac{1}{2}+ir_j)$ are defined in (5.6).

Proof. The statements (1) and (3) follow from the combination of Proposition 5.1 and Proposition 5.3. The case $\sigma = X_+\phi_{ir_k}$ follows from Proposition 2.3 and the fact that the Patterson-Sullivan distributions are invariant under time-reversal (cf. Proposition 3.3). More precisely, by Theorem 1.1

$$\langle Op(X_+\phi_{ir_k})\phi_{ir_j}, \phi_{ir_j}\rangle = 2^{(1+2ir_j)} \int_{S\mathbf{D}} (L_{r_j}\chi X_+\phi_{ir_k})(g) PS_{ir_j}(dg),$$

and by (1.7)

$$\int_{S\mathbf{D}} (L_{r_j}\chi X_+\phi_{ir_k}) PS_{ir_j}(dg) = \int_{\mathbf{R}} (1+u^2)^{-(\frac{1}{2}+ir_j)} I_{PS_{ir_j}}(X_+\phi_{ir_k})(u)du,$$

with

$$I_{PS_{ir_j}}(X_+\phi_{ir_k})(u) := \int_{\Gamma\backslash G} X_+\phi_{ir_k}(gn_u) PS_{ir_j}(dg).$$

But $X_+\phi_{ir_k}(gn_u) = \frac{d}{du}\phi_{ir_k}(gn_u)$ and after integrating by parts we have

$$\langle Op(X_+\phi_{ir_k})\phi_{ir_j}, \phi_{ir_j}\rangle = 2^{(1+2ir_j)}(\frac{1}{2}+ir_j) \int_{\mathbf{R}} (1+u^2)^{-(\frac{3}{2}+ir_j)}(2u) I_{PS_{ir_j}}(\phi_{ir_k})(u)du. \quad (6.17)$$

By Proposition 5.1 with $m = 0$, and the weight zero calculation in (5.7), we see that the even F term makes no contribution to (6.17) since it is the integral of an odd function times an even function. Hence, only the odd G term contributes and we see that $\langle Op(X_+\phi_{ir_k})\phi_{ir_j}, \phi_{ir_j}\rangle$ is a constant multiple of $\langle X_+\phi_{ir_k}, PS_{ir_j}\rangle$. But this vanishes since $X_+\phi_{ir_k}$ is odd under time reversal while PS_{ir_j} is even.

□

We note that these explicit formula give a new proof of Theorem 1.1:

Corollary 6.7. *When σ is a joint (Ω, W)-eigenfunction, we find again that $\langle Op(\sigma)\phi_{ir}, \phi_{ir}\rangle$ is asymptotically the same as $r^{-1/2}\langle \sigma, PS_{ir}\rangle_{S\mathbf{X}_\Gamma}$.*

Proof. By definition, $F_{\tau,m}(1/2) = 1$ whereas $G_{\tau,m}(1/2) = 0, G'(1/2) = -2i$. The stationary phase method then shows that

$$\int (1+u^2)^{-(\frac{1}{2}+ir)} F_{\tau,m}\left(\frac{u-i}{-2i}\right) du \sim r^{-1/2}$$

whereas

$$\int (1+u^2)^{-(\frac{1}{2}+ir)} G_{\tau,m}\left(\frac{u-i}{-2i}\right) du \sim r^{-3/2}.$$

Here, we use the estimates in (5.8), which can be generalized in all weights.

□

7. Dynamical zeta-functions: Thermodynamic formalism

In this part,we prove Theorem 1.3 for $\mathcal{Z}_2$, showing that it has a meromorphic continuation to $\mathbf{C}$, identifying the poles in the strip $0 < \Re e(s) < 1$ and the residues. We use the thermodynamic formalism introduced by Ruelle [Ru87] to study the "resonances" of the geodesic flow.

Let us make a short digression to describe certain aspects of Ruelle's work [Ru87]. His aim was to study the Fourier transform of the correlation function,

$$\rho_{F,G}(t) = \int F(x)G(g^t x)d\omega(x) - \int F d\omega \int G d\omega,$$

$(t \geq 0)$, in the very general context of an Axiom A flow (g^t) (for instance, when ω is the measure of maximal entropy). He showed, for smooth enough functions F and G, that the Fourier transform $\hat{\rho}_{F,G}$ has a meromorphic extension to a half-plane of the form $\{\Re e(s) > h - \varepsilon\}$, strictly beyond its half-plane of absolute convergence $\{\Re e\ s > h\}$ (where h represents, in a general context, the topological entropy). He used the so-called "thermodynamic formalism" and showed that the poles of $\hat{\rho}(s)$ occurred precisely for certain values s, linked with the existence of distributions obeying specific transformation rules.

In the case of the geodesic flow on a compact surface of constant curvature -1, and for C^1 functions F, G on $\Gamma\backslash G$, the Fourier transform $\hat{\rho}$ is an analytic function in the half-plane $\{s, \Re es > 1\}$ and has a meromorphic extension to $\{\Re es > 0\}$ with poles at values of $s = \frac{1}{2} + ir$ for which there exists a distribution e_{ir} on $S\mathbf{X}_\Gamma$ satisfying:

- $g^t.e_{ir} = e^{-(\frac{1}{2}+ir)t} e_{ir}$
- e_{ir} is invariant under the stable horocycle flow.

In the case of constant negative curvature, it follows that e_{ir} is given by:

$$\langle F, e_{ir} \rangle_{S\mathbf{X}_\Gamma} = \int F(z,b) e^{(\frac{1}{2}+ir)\langle z,b \rangle} T_{ir}(db) dVol(z) = \int_{\mathbf{X}_\Gamma} Op(F)\phi_{ir}(z) dVol(z) \tag{7.1}$$

where T_{ir} is the boundary values of an eigenfunction ϕ_{ir} of $\triangle$ of eigenvalue $\frac{1}{4} + r^2$ (see equation (3.5), and [Z2]) Hence the poles of $\hat{\rho}$, i.e. the Ruelle resonances, occur at $s_n = 1/2 + ir_n$. If the eigenvalue is simple, the residue of $\hat{\rho}_{a,b}$ at s_n is given, up to multiplicative normalization, by

$$\begin{aligned} R_{r_n}(F,G) &= \left(\int F(z,b) e^{(1/2+ir_n)\langle z,b \rangle} T_{ir_n}(db) dVol(z) \right) \\ &\quad \times \left(\int G \circ \iota(z,b) e^{(1/2+ir_n)\langle z,b \rangle} T_{ir_n}(db) dVol(z) \right) \\ &= \left(\int_{\mathbf{X}_\Gamma} Op(F)\phi_{ir_n} dVol(z) \right) \left(\int_{\mathbf{X}_\Gamma} Op(G \circ \iota)\phi_{ir_n} dVol(z) \right) \\ &= \langle F, e_{ir_n} \rangle_{S\mathbf{X}_\Gamma} \langle G \circ \iota, e_{ir_n} \rangle_{S\mathbf{X}_\Gamma}, \end{aligned}$$

where ι denotes time reversal. To see this, we observe that the residue $R_{r_n}(F,G)$ is bilinear in F and G, and its definition implies that it satisfies the identities

$$R_{r_n}(F \circ g^t, G) = R_{r_n}(F, G \circ g^{-t}) = e^{-(1/2+ir_n)t} R_{r_n}(F,G),$$

and

$$R_{r_n}(F \circ h^u_+, G) = R_{r_n}(F, G \circ h^u_-) = R_{r_n}(F,G),$$

where h^u_+ denotes the stable horocyclic flow and h^u_- the unstable one. It follows that it must equal $\langle F, e_{ir_n}\rangle_{S\mathbf{X}_\Gamma} \langle G \circ \iota, e_{ir_n}\rangle_{S\mathbf{X}_\Gamma}$ if the eigenvalue $1/4 + r_n^2$ is simple. If the eigenvalue is not simple, the expression becomes more complicated, as one has to form a linear combination of the functionals associated to the various eigenfunctions.

In the same spirit, we now prove Theorem 1.3 concerning the meromorphic continuation of $\mathcal{Z}_2$. We use the methods developed by Rugh [Rugh92, Rugh96] in real-analytic situations.

Remark 7.1. Although the poles of $\mathcal{Z}_2$ will turn out to be the same as those of $\hat{\rho}$ (the Ruelle resonances), the residues cannot be the same: the residues of $\mathcal{Z}_2$ must define geodesic flow invariant distributions, whereas the residues of $\hat{\rho}$ define horocyclic invariant distributions as explained above.

7.1. Markovian coding of the boundary.

The proof given here relates the function $\mathcal{Z}_2$ to the determinant of certain operators, called transfer operators. To define them, we need to recall from [Se] the construction of Markov sections, using the Bowen-Series coding of the action of Γ on the boundary B. Series used this construction to study Poincaré series. We apply it to the somewhat different objects $\mathcal{Z}_2$. For this application, we need some further discussion of Markov coding which we could not find in the literature.

If we want to study the action of Γ on the boundary, and the existence of conformally invariant distributions – by this, we mean the property 3.4 – it is enough to consider a set of generators of Γ. In fact, it is even enough to work with a single, well chosen transformation of the boundary: roughly speaking, this transformation $F^{(r)}$ will be defined by cutting the boundary B into a finite number of closed intervals J_i, and will act on each J_i by the action of one of the chosen generators of Γ.

We will require the map $F^{(r)} : J = \sqcup J_i \longrightarrow J$ to have the following properties:

(i) $F^{(r)}$ is analytic, expanding (or at least, some power of $F^{(r)}$ is expanding).

(ii) The J_is form a Markov partition for $F^{(r)}$. This means that $F^{(r)}$ sends the boundary of J to itself.

(iii) The natural map $J = \sqcup J_i \longrightarrow B$ gives a bijection between periodic points of $F^{(r)}$ and points at infinity of closed geodesics, except for the closed geodesics ending on the boundary of an interval J_i, that have exactly two preimages. If ${F^{(r)}}^n x = x$, and γ is the closed geodesic corresponding to x, then $|({F^{(r)}}^n)' x| = e^{L_\gamma}$.

We recall briefly the construction of $F^{(r)}$ proposed by Series [Se], when Γ is cocompact: she chooses a symmetric generating set for Γ, called Γ_0. She then defines a notion of "admissible representation" of an element $g \in \Gamma$ as a word $g = g_1 ... g_n$ with $g_j \in \Gamma_0$, such that

– an admissible word is a shortest representation of g in the alphabet Γ_0.

– every $g \in \Gamma_0$ has a unique admissible representation.

Without going into details, admissible words are shortest word-representations; and besides, whenever there is a choice of several such representations, one chooses the one that "turns the furthest possible to the right" in the Cayley graph of Γ with respect to Γ_0 (seen as a subset of the hyperbolic plane).

Let us denote $\Sigma_f^{(r)}$ the set of finite admissible words; the notation is borrowed from [Se] but we are adding an r to specify that we are choosing representations that turn the most possible right in the Cayley graph – the same convention Series used in her paper. Replacing "right" by "left" one would obtain another notion of admissible words, and we denote $\Sigma_f^{(l)}$ the set of left-admissible words. Note that $\Sigma_f^{(l)} = {\Sigma_f^{(r)}}^{-1}$. Now define $\Sigma^{(r)}$, the set of infinite right-admissible words, as

$$\Sigma^{(r)} = \{(g_j) \in \Gamma_0^{\mathbf{Z}_+}, g_j ... g_{j+k} \in \Sigma_f^{(r)}, \forall j, k \geq 0\}.$$

Series shows in [Se] that the map

$$\Sigma_f^{(r)} \longrightarrow \mathbf{H}, \quad g_1 ... g_n \mapsto g_1 ... g_n.0 \tag{7.2}$$

can be extended to a continuous map $j^{(r)} : \Sigma^{(r)} \longrightarrow B$. She denotes $I^{(r)}(g_j)$ the set of points in B that have a representation by a sequence in $\Sigma^{(r)}$ starting with the generator g_j. The boundary B is thus divided into a finite number of closed intervals, with disjoint interiors. One can define a map $F^{(r)}$ that acts on $\Sigma^{(r)}$ by deleting the first symbol and shifting the sequence to the left; seen as a map on B, it acts as g_j^{-1} on each interval $I^{(r)}(g_j)$. Actually, the map $F^{(r)}$ is defined on $I^{(r)} := \sqcup I^{(r)}(g_j)$; when working on B one should always remember that its definition is ambivalent on boundary points. The partition $B = \cup I^{(r)}(g_j)$ is not exactly a Markov partition for the action of $F^{(r)}$: there is no reason that boundary points should be sent to boundary points. But, by construction, the images of these boundary points under iteration of $F^{(r)}$ form a finite set. Cutting the intervals $I^{(r)}(g_j)$ at these points, one can refine the partition $B = \cup I^{(r)}(g_j)$ into a new finite partition $B = \cup J_j$ that is now Markov. This way we obtain a transformation $F^{(r)}$ satisfying all the conditions (i), (ii), (iii). An element of B may be coded either by a word in $\Sigma^{(r)}$, as we have already seen, or by an element of the subshift of finite type

$$\Lambda^{(r)} = \{(i_k)_{k \geq 0}, F^{(r)}(J_{i_k}) \cap int(J_{i_{k+1}}) \neq \emptyset \text{ for all } k \geq 0\}.$$

Both codings are bijective except on a countable set (in fact the coding map is at most 2 to 1).

To make the link with the geodesic flow, we now extend the expanding transformation $F^{(r)}$ to an invertible transformation F of a subset of $B \times B$; in terms of symbolic dynamics, we want to work with two-sided subshifts. We consider

$$\Sigma^{(l)} = \{(g_j) \in \Gamma_0^{\mathbf{Z}_+^*}, g_j...g_{j+k} \in \Sigma_f^{(l)}, \forall j, k > 0\}$$

and

$$\Sigma_{(l)} = \{(g_j) \in \Gamma_0^{\mathbf{Z}_-^*}, g_{j-k}...g_j \in \Sigma_f^{(r)} \forall j, k < 0\}.$$

Formally, elements of $\Sigma_{(l)}$ are inverses of elements of $\Sigma^{(l)}$. By the same consideration as before, we have a coding map $j^{(l)} : \Sigma^{(l)} \longrightarrow B$ or equivalently $j_{(l)} : \Sigma_{(l)} \longrightarrow B$; we denote $I_{(l)}(g_j) \subset B$ the interval formed by points whose coding in $\Sigma_{(l)}$ ends with g_j. This gives a partition $B = \cup I_{(l)}(g_j)$ and a map $F_{(l)}$ on B, that corresponds to the right-shift on $\Sigma_{(l)}$. We can refine the partition $B = \cup I_{(l)}(g_j)$ into a Markov partition $B = \cup K_j$ and code the dynamics by a one-sided subshift of finite type $\Lambda_{(l)}$.

Let us now introduce the two-sided subshift Σ,

$$\Sigma = \{(g_j) \in \Gamma_0^{\mathbf{Z}}, g_j...g_{j+k} \in \Sigma_f^{(r)}, \forall j, k\};$$

Σ is in natural bijection with a subset of $\Sigma_{(l)} \times \Sigma^{(r)}$, and thus there is a coding map from Σ to a subset $X \subset B \times B$:

$$\begin{array}{rcl} j : \Sigma & \longrightarrow & X, \\ j(\sigma_{(l)}, \sigma^{(r)}) & = & \Big(j_{(l)}(\sigma_{(l)}), j^{(r)}(\sigma^{(r)})\Big) \end{array}$$

The shift to the left on Σ gives an invertible transformation F on X; note, as above, that F is actually well defined on a subset of $\sqcup I_{(l)}(g_j) \times \sqcup I^{(r)}(g_k)$ and is defined ambivalently at certain points of X. If y is in $I^{(r)}(g_j)$ then $F(x,y) = (g_j^{-1}x, g_j^{-1}y)$, or in other words $F(x,y) = (G_{(l)_j}x, F^{(r)}y)$, where $G_{(l)_j}$ is the inverse branch of $F_{(l)}$ taking values in $I_{(l)}(g_j)$.

The partition of X into $\sqcup I_{(l)}(g_j) \times \sqcup I^{(r)}(g_k)$ is not a Markov partition for the action of F, but $X = \cup(K_j \times J_i)$ is Markov. The action of F is then semiconjugate to the left-shift on the subshift of finite type

$$\Lambda = \{(m_k, n_k)_{k\in\mathbf{Z}} / F(K_{m_k} \times J_{n_k}) \cap int(K_{m_{k+1}} \times J_{n_{k+1}}) \neq \emptyset \text{ for all } k \in \mathbf{Z}\}.$$

We can identify $X \subset B \times B$ with a transversal for the action of the geodesic flow: we observe that, for each (x,y) in X, the geodesic $\gamma_{x,y}$ contains a unique vector, denoted $v_{x,y}$, which is in the stable manifold of a vector based at 0. To recover the action of the geodesic flow on the whole tangent bundle, one needs to add a time parameter measuring the time it takes to go from (x,y) to $F(x,y)$. Because $\gamma_{x,y}$ and $\gamma_{F(x,y)}$ represent the same geodesic in the quotient $\Gamma\backslash G$, there exists $\tau(x,y) \in \mathbf{R}$ such that $g^\tau v_{x,y} = v_{F(x,y)}$. More precisely, the function τ is defined without any ambiguity on $\sqcup \overset{\circ}{K}_j \times \overset{\circ}{J}_i$ and can be extended to a continuous function on $\sqcup K_j \times J_i$. By construction, the function τ is locally constant on stable manifolds; i.e., it depends only on the variable y. It is analytic on each rectangle $K_j \times J_i$. We see τ as a return time from the section X to itself; note however that τ may change sign: we are not exactly in the usual situation of a "first return time". Nevertheless, when y is a periodic point of period n of $F^{(r)}$, and γ is the corresponding closed geodesic, we have $\sum_{k=0}^{n-1} \tau({F^{(r)}}^k y) = L_\gamma > 0$.

We have a surjection, almost one-to-one, from

$$X^\tau := \{((x,y),t) \in X \times \mathbf{R}, t \in [0, \tau(x)]\}$$

to the unit tangent bundle, defined by saying that the image of $((x,y),s)$ is $g^s v_{x,y}$, the image of $v_{x,y}$ under time s of the geodesic flow. This surjection is not one-to-one on boundary points; by definition of τ, $((x,y),\tau(y))$ has the same image as $(F(x,y),0)$. On X^τ the geodesic flow corresponds the translation of the parameter t.

7.2. Transfer operators.

Let us first recall briefly how the main results of [Rugh92] or [Rugh96] read in our context. We follow very closely the notations of these papers.

Consider two rectangles $\Omega = K \times J$ and $\Omega' = K' \times J'$ such that $F(K \times J) \cap \mathit{int}\ K' \times J' \neq \emptyset$. This means that there exists an element g in the generating set Γ_0 such that $g^{-1}K \subset K'$ and $gJ' \subset J$. The maps g and g^{-1} are Moebius transformations, hence analytic. Obviously, we can find complex simply connected compact neighbourhoods $\mathcal{D}(K)$, $\mathcal{D}(J)$, $\mathcal{D}(K')$, $\mathcal{D}(J') \subset \mathbf{C}$ with $K \subset \mathit{int}\,\mathcal{D}(K)$, $J \subset \mathit{int}\,\mathcal{D}(J)$, $K' \subset \mathit{int}\,\mathcal{D}(K')$, $J' \subset \mathit{int}\,\mathcal{D}(J')$ such that $g^{-1}\overline{\mathcal{D}}(K) \subset \mathit{int}\,\mathcal{D}(K')$ and $g\overline{\mathcal{D}}(J') \subset \mathit{int}\,\mathcal{D}(J)$. In the terminology of [Rugh92], we obtain a *real analytic hyperbolic map* f between the rectangles Ω and Ω' (with complex rectangles $\Omega_{\mathbf{C}} = \mathcal{D}(K) \times \mathcal{D}(J)$, $\Omega'_{\mathbf{C}} = \mathcal{D}(K') \times \mathcal{D}(J')$) by letting $f(w_1,w_2) = (z_1,z_2) = (g^{-1}w_1, g^{-1}w_2)$. In this context, the maps called ϕ_s^1 and ϕ_u^1 by Rugh are very simple, they depend only on one coordinate : $\phi_s^1(w_1,z_2) = gz_2$, $\phi_u^1(w_1,z_2) = g^{-1}w_1$.

Always following [Rugh92], we introduce the Banach space U_Ω of functions which are analytic in $(\overline{\mathbf{C}} \setminus \mathcal{D}(K)) \times \mathcal{D}(J)$ with a continuous extension to the boundary (endowed with the sup norm); $U_{\Omega'}$ is defined similarly. The transfer operator associated to the two rectangles Ω, Ω' sends $\chi \in U_\Omega$ to an element of $U_{\Omega'}$, as follows :

$$L_{\Omega,\Omega'}\chi(z_1,z_2) = \int_{\partial\mathcal{D}(K),\partial\mathcal{D}(J)} \frac{dw_1}{2\pi i}\frac{dw_2}{2\pi i}\,\frac{1}{z_1 - g^{-1}w_1}\,\frac{g'(z_2)}{w_2 - g.z_2}\chi(w_1,w_2).$$

In other words, for every function ψ analytic in $\mathcal{D}(K')$ and continuous on the boundary (which we denote $\psi \in Hol(\mathcal{D}(K'))$), we have

$$\int_{\partial\mathcal{D}(K')} L_{\Omega,\Omega'}\chi(z_1,z_2)\psi(z_1)\frac{dz_1}{2\pi i} = \int_{\partial\mathcal{D}(K)} \frac{dw_1}{2\pi i} g'(z_2)\psi(g^{-1}w_1)\chi(w_1,gz_2)\frac{dw_1}{2\pi i}.$$

The full transfer operator is obtained by considering all possible pairs of rectangles (Ω,Ω'). Because $F^{(r)}$ is eventually expanding and the inverse branches $G_{(l)}$ are contracting, it is possible to choose complex discs $\mathcal{D}(K_i)$, $\mathcal{D}(J_j)$ which are suitable for all pairs $(\Omega = K_j \times J_i, \Omega' = K_{j'} \times J_{i'})$. The transfer operator acts on $U = \oplus U_\Omega$, and is defined by $(L\chi)_{\Omega'} = \sum_{\Omega} L_{\Omega,\Omega'}\chi_{\Omega}$. We have

$$\int_{\partial\mathcal{D}(K)} L\chi(z_1,z_2)\psi_K(z_1)\frac{dz_1}{2\pi i} = \sum_{k_{-1},k_0} g'_{k_0}(z_2)\int_{\partial\mathcal{D}(K_{k_{-1}})}\psi_K(g_{k_0}^{-1}w_1)\chi(w_1,g_{k_0}z_2)\frac{dw_1}{2\pi i}$$

for every $\psi = (\psi_{K_j}) \in \oplus_j Hol(\mathcal{D}(K_j))$. The sum runs over all k_{-1}, k_0 such that the rectangle $K_{k_{-1}} \times J_{k_0}$ contains a point (w_1,w_2) with $F(w_1,w_2) = (z_1,z_2)$. The notation g_{k_0} means the element of Γ_0 such that $F(w_1,w_2) = (g_{k_0}^{-1}w_1, g_{k_0}^{-1}w_2)$ if $(w_1,w_2) \in K_{k_{-1}} \times J_{k_0}$. Rugh shows that this operator is nuclear and that

$$Tr(L^n) = \sum_{F^n w = w}\frac{1}{|det(DF^n_w - 1)|}.$$

Note that a fixed point $F^n w = w$ corresponds to a closed geodesic represented by $\gamma \in \Gamma$; and $|det(DF^n_w - 1)| = |(\gamma'_w - 1)(\gamma_w^{-1\prime} - 1)| = |(e^{L_\gamma} - 1)(e^{-L_\gamma} - 1)| = 4\sinh(L_\gamma/2)^2$.

For our purposes we need a variant of Rugh's transfer operators. Let a be an analytic function on $S\mathbf{X}_\Gamma$. Let A be the real-analytic function on $X = \sqcup(K_j \times J_i)$, defined by

$$A(w_1,w_2) = \int_0^{\tau(w_2)} a(w_1,w_2,t)\,dt$$

if $(w_1,w_2) \in X \subset B \times B$. In other words, A is the Radon transform $A = \mathcal{R}(\chi a)$ defined in (3.11), and χ is the cut-off function $\chi((w_1,w_2),s) := \sum_{i,j}\mathbf{1}_{J_i}(w_2)\mathbf{1}_{K_j}(w_1)$ $\mathbf{1}_{(0,\tau(w_2))}(s)$. If $w = (w_1,w_2)$ is a periodic point of period n for F, and if γ is the corresponding closed geodesic, then

$$S_n A(w) := \sum_{k=0}^{n-1} A(F^k w) = \int_\gamma a.$$

We introduce a family of transfer operators $L_{s,z}$ $(s,z \in \mathbf{C})$, acting on the Banach space U defined above:

$$L_{s,z}\chi(z_1,z_2) = \sum_{k_{-1},k_0}\int_{\partial\mathcal{D}(K_{k_{-1}}),\partial\mathcal{D}(J_{k_0})}\frac{dw_1}{2\pi i}\frac{dw_2}{2\pi i}\frac{(g_{k_0}^{-1\prime}(w_1))^{s/2}}{z_1 - g_{k_0}^{-1}w_1}\frac{(g'_{k_0}(z_2))^{s/2}}{w_2 - g_{k_0}.z_2}e^{zA(w_1,w_2)}\chi(w_1,w_2).$$

In other words, we have

$$\int_{\partial\mathcal{D}(K)} L_{s,z}\chi(z_1,z_2)\psi(z_1)\frac{dz_1}{2\pi i} = \sum_{k_{-1},k_0} g'_{k_0}(z_2)^{1+(s/2-1)} \int_{\partial\mathcal{D}(K_{k_{-1}})}\psi(g_{k_0}^{-1}w_1)\chi(w_1,g_{k_0}z_2)e^{zA(w_1,g_{k_0}z_2)}g_{k_0}^{-1\prime}(w_1)^{s/2}\frac{dw_1}{2\pi i} \tag{7.3}$$

for every $\psi = (\psi_{K_j}) \in \oplus_j Hol(\mathcal{D}(K_j))$.

The operators $L_{s,z}$ are bounded on U (they are even nuclear) and depend analytically on (s, z), as is easily seen in the expression above.

We caution that the notation $L_{s,z}$ has nothing to do with the operators L_r used in the previous sections: this should cause no confusion, as this section is rather disjoint from the others.

7.3. Determinants and zeta functions.

Apart from the introduction of the weight A, our transfer operator also differs from Rugh's by the terms $g'_{k_0}(z_2)^{(s/2-1)}$ and $g_{k_0}^{-1\prime}(w_1)^{s/2}$. All his arguments can be adapted with obvious modifications to this situation, and we do not reproduce them here. In paragraph 4.4 of [Rugh92], it is shown that $L_{s,z}$ is a nuclear (trace class) operator. One can take the determinant of $I - L_{s,z}$:

$$d(s, z) := det(I - L_{s,z}) = \prod (1 - \lambda_{s,z}^{(i)})^{m^{(i)}}$$

where the product runs over the spectrum of $L_{s,z}$, and $m^{(i)} = m_{s,z}^{(i)}$ denotes the multiplicity of $\lambda^{(i)}$. The eigenvalues do not necessarily depend analytically on (s, z), as the multiplicity may vary; the determinant $d(s, z)$, however, is an analytic function of (s, z):

For given (s_0, z_0), consider, for every i, a neighbourhood V_i of $\lambda_{s_0,z_0}^{(i)}$, such that the V_is are all pairwise disjoint. Let $P_{s,z}^i$ be the spectral projector on V_i for the operator $L_{s,z}$: $P_{s,z}^i$ depends analytically on (s, z), in a neighbourhood of (s_0, z_0). Call $B_{s,z}^i = L_{s,z} P_{s,z}^{(i)}$: these are operators of rank $m_{s_0,z_0}^{(i)}$, depending analytically on (s, z) in a neighbourhood of (s_0, z_0). By definition the spectrum of $B^{(i)}(s, z)$ is contained in V_i. Of course,

$$det\,(1 - B_{s_0,z_0}^{(i)}) = (1 - \lambda_{s_0,z_0}^{(i)})^{m_{s_0,z_0}^{(i)}}.$$

One can write, in the neighourhood of (s_0, z_0),

$$d(s, z) = \prod_i det\,(1 - B_{s,z}^{(i)}) \tag{7.4}$$

This shows that the determinant $d(s, z)$ is an entire function, and has zeros exactly when $L_{s,z}$ has the eigenvalue 1.

Let us introduce the notations $\tau_1(w) = -\log g_{k_0}^{-1\prime}(w_1)$ and $\tau_2(w) = \log g'_{k_0}(w_2)$ if $w = (w_1, w_2) \in X$ with $w_2 \in J_{k_0}$. Rugh shows that the following remarkable identities hold; For all n, the trace of $L_{s,z}^n$ is

$$Tr(L_{s,z}^n) = \sum_{w, F^n w = w} \frac{e^{zS_n A(w) - sS_n\tau_1(w) + (1-s/2)S_n\tau_2(w)}}{|det(DF^n(w) - 1)|}. \tag{7.5}$$

It follows that

Proposition 7.1.

$$d(s,z) := det(I - L_{s,z}) = \exp\left(-\sum_{w,F^n w=w} \frac{1}{n}\frac{e^{zS_nA(w)-sS_n\tau_1(w)+(1-s/2)S_n\tau_2(w)}}{|det(DF^n(w)-1)|}\right). \tag{7.6}$$

In particular, the function

$$\frac{\partial_z d}{d}(s,0) = -\sum_{w,F^n w=w}\frac{1}{n}\frac{S_nA(w)}{det|DF^n(w)-1|}$$

has poles exactly when 1 is in the spectrum of $L_{s,0}$.

Because periodic points of F correspond to closed geodesics, we can express (7.6) in terms of periodic geodesics. If $F^n w = w$ and γ is the corresponding closed geodesics, we have $S_nA(w) = \int_\gamma a$, $S_n\tau_1(w) = S_n\tau_2(w) = L_\gamma$. Thus, $d(s,z)$ is more or less the same as

$$\exp\left(-{\sum_\gamma}' \sum_{p\geq 1}\frac{1}{p}\frac{e^{p(z\int_\gamma a-(s-1)L_\gamma)}}{|\sinh(pL_\gamma/2)|^2}\right) = \prod_\gamma \prod_{m,n\in\mathbf{N}}(1 - e^{z\int_\gamma a-(s+m+n)L_\gamma}). \tag{7.7}$$

A “prime” following a sum or a product means we are summing over primitive closed orbits. Otherwise, we sum or take the product over all closed geodesics.

The previous formula, however, is not exactly true, because certain periodic geodesics correspond to several different periodic orbits of F; namely, those going through the boundary of X (there are a finite number of them). The precise expression of $d(s,z)$ in terms of closed geodesics is given in [Rugh96], or [Mo]:

$$d(s,z) = \prod_\gamma \prod_{m,n\in\mathbf{N}}(1 - e^{z\int_\gamma a-(s+m+n)L_\gamma})\ .P(s,z) \tag{7.8}$$

where the correction term is

$$P(s,z) = \frac{\prod_c \prod_{m,n\geq 0}(1 - e^{-(s+m+n)l(c)+z\int_c a})}{\prod_{c'} \prod_{m,n\geq 0}(1 - e^{-(s+m+n)l(c')+z\int_{c'} a})}; \tag{7.9}$$

the products run over a finite number of periodic orbits that are counted several times in the Markov coding. The correction factor on the right is analytic and non-vanishing in $\{\Re e\ s > 0\}$, thus the zeros of the two functions (7.6) and (7.7) are the same there.

Remark 7.2. In the half-plane $\{\Re e\ s \leq 0\}$, the correction factor $P(s,z)$ is more difficult to analyze because it seems that its singularities could depend on the choice of the Markov section X. It was, however shown in [Rugh96] that the apparent singularities of (7.7), arising from the identity (7.8), are removable.

Remark 7.3. For $z = 0$ (which is the case treated in [Rugh96]) we obtain the relation

$$d(s,0) = \prod_{n\in\mathbf{N}} \zeta_S(s+n)\ .P(s,0) \tag{7.10}$$

where ζ_S is the Selberg zeta function. In particular, $d(s,0)$ has the same singularities as ζ_S in $\{\Re e\ s > 0\}$.

We focus our attention in the region $\{\Re e\ s > 0\}$. There, the function $\frac{\partial_z d}{d}(s,0)$ has the same singularities as

$$\mathcal{Z}_2(s,0) := \sum_\gamma \left(\int_{\gamma_0} a\right) \frac{e^{-(s-1)L_\gamma}}{|\sinh(L_\gamma/2)|^2}.$$

This shows that the singularities of $\mathcal{Z}_2$ appear when $L_{s,0}$ has 1 as an eigenvalue. In the next paragraph, we show that this occurs for $s = \frac{1}{2} \pm ir_n$. Then we identify the residues.

Remark 7.4. In our conventions, $r_n \geq 0$ and we have defined the boundary values T_{ir_n} using this choice of sign. For simplicity, we will restrict our attention to $s = s_n = \frac{1}{2} + ir_n$, but the analysis at $s = (1 - s_n) = \frac{1}{2} - ir_n$ would be similar.

7.3.1. Location of poles of $\mathcal{Z}_2$ in the critical strip. For $s = s_n$, one can check directly that 1 is in the spectrum of $L_{s,0}$: the eigenspace is spanned by the functionals

$$\chi_{(l)}^{ir_n}(z_1, z_2) = \frac{T_{ir_n}(dz_1)}{|z_1 - z_2|^{s_n}}, \tag{7.11}$$

where T_{ir_n} are the boundary values of eigenfunctions of the Laplacian ($(z_1, z_2) \in X = \sqcup(K_j \times J_i)$). The functionals $\chi_{(l)}^{ir_n}(z_1, z_2)$ are analytic with respect to z_2 and are distributions of order 1 with respect to z_1 : in particular, they belong to the Banach space U. If $\psi = (\psi_{K_j}) \in \oplus_j Hol(\mathcal{D}(K_j))$, it defines, of course, a C^∞ function on each interval K_j. For $\chi = \chi_{(l)}^{ir_n}$, the integral on the right-hand side of (7.3) is nothing but the pairing between the distribution $L_{s_n,0}\chi_{(l)}^{ir_n}$ and the function ψ. Identity (7.3) can be extended to $\psi \in \oplus_j C^\infty(K_j)$ (or even $\psi \in \oplus_j C^1(K_j)$, since we know T_{ir_n} is of order 1). Integrals should now be understood as the pairing between distributions and C^∞ functions.

To show that $L_{s_n,0}\chi_{(l)}^{ir_n} = \chi_{(l)}^{ir_n}$, we need to check that

$$\int_K \psi(z_1) \frac{T_{ir_n}(dz_1)}{|z_1 - z_2|^{s_n}} = \sum_{k_{-1},k_0} g_{k_0}'(z_2)^{s_n/2} \int_{K_{k_{-1}}} \psi(g_{k_0}^{-1} w_1) g_{k_0}^{-1\prime}(w_1)^{s_n/2} \frac{T_{ir_n}(dw_1)}{|w_1 - g_{k_0} z_2|^{s_n}},$$

for any $\psi \in \oplus_j C^\infty(K_j)$. Again, the integrals should be understood as a notation for the pairing between distributions and functions. If ψ is analytic, $\int_K$ is given, as in the previous paragraph, by the integral on the contour $\partial\mathcal{D}(K)$.

Using the invariance properties of T_{ir_n}, the last expression can be transformed to

$$\sum_{k_{-1},k_0} g'_{k_0}(z_2)^{1+s_n} \int_{g_{k_0}^{-1}K_{k_{-1}}} \psi(z_1) g_{k_0}^{-1\prime}(g_{k_0}z_1)^{s_n/2} \frac{T_{ir_n}(dg_{k_0}z_1)}{|g_{k_0}z_1 - g_{k_0}z_2|^{s_n}}$$

$$= \sum_{k_{-1},k_0} g'_{k_0}(z_2)^{s_n/2} \int_{g_{k_0}^{-1}K_{k_{-1}}} \psi(z_1) g_{k_0}^{-1\prime}(g_{k_0}z_1)^{s_n/2} g'_{k_0}(z_1)^{s_n} \frac{T_{ir_n}(dz_1)}{|z_1 - z_2|^{s_n}}$$

$$g_{k_0}^{-1\prime}(z_1)^{-s_n/2} g_{k_0}^{-1\prime}(z_2)^{-s_n/2} = \int \psi(z_1) \frac{T_{ir_n}(dz_1)}{|z_1 - z_2|^{s_n}},$$

which is the desired property.

Remark 7.5. Similarly, the functionals

$$\chi_{(r)}^{ir_n}(z_1, z_2) = \frac{T_{ir_n}(dz_2)}{|z_1 - z_2|^{s_n}} \tag{7.12}$$

are eigenvectors for the adjoint $L^*_{s_n,0}$.

Conversely, we need to know that 1 is in the spectrum of $L_{s,0}$ only if s is one of the s_n; and that the multiplicity of 1 is exactly the multiplicity of $s_n(1-s_n)$ in the spectrum of the Laplacian (this means that $L_{s_n,0}$ has no Jordan block associated to the eigenvalue 1). We can see no direct way of proving this last fact without using the relation with the Selberg zeta-function (Remark 7.3). For the latter we know indeed that the zeros occur when $s(1-s)$ is in the spectrum of the Laplacian, with the same multiplicity.

7.3.2. The residues. We are interested in the singularities of $\mathcal{Z}_2$ in $\{\Re e\ s > 0\}$, or equivalently in the singularities of

$$\frac{\partial_z d(s,0)}{d(s,0)} = \sum_i \frac{\partial_z det\,(1 - B_{s,0}^{(i)})}{det\,(1 - B_{s,0}^{(i)})} \tag{7.13}$$

From the previous paragraph, we know that $d(s,0) = 0$ if and only if $s = \frac{1}{2} \pm ir_n$ (where $1/4 + r_n^2$ is an eigenvalue of the Laplacian). For some i, the operator $B_{s_n,0}^{(i)}$ has 1 as an eigenvalue, and its multiplicity m_i is the same as the multiplicity of $1/4 + r_n^2$ in the spectrum of the Laplacian. As in the previous paragraph, we treat the case of $s = s_n = \frac{1}{2} + ir_n$; the case of $s = \frac{1}{2} - ir_n$ would be similar except for the choice of a different convention in the definition of boundary values. We will see that the singularity of the function (7.13) at $s = s_n$ is a pole; the residue must then be given by

$$m_i \frac{\partial_z \partial_s^{m_i - 1} det\,(1 - B_{s_n,0}^{(i)})}{\partial_s^{m_i} det\,(1 - B_{s_n,0}^{(i)})}.$$

Theorem 1.3 will then follow directly from:

Proposition 7.2.

$$m_i \frac{\partial_z \partial_s^{m-1} det\,(1 - B^{(i)}_{s_n,0})}{\partial_s^m det\,(1 - B^{(i)}_{s_n,0})} = \sum_{r_j = r_n} \frac{\langle \chi a, PS_{ir_j} \rangle_{S\mathbf{D}}}{\langle \chi, PS_{ir_j} \rangle_{S\mathbf{D}}}.$$

Proof. If $1/4 + r_n^2$ is an eigenvalue of the Laplacian of multiplicity m, we know that 1 is an eigenvalue of $L_{s_n,0}$ of multiplicity m. We also know – and this is rather important – that the eigenvalue 1 corresponds to a diagonal block for $L_{s_n,0}$.

Let $V \subset \mathbf{C}$ be a neighbourhood of 1 that does not meet the rest of the spectrum of $L_{s_n,0}$. Let $P_{s,z}$ be the spectral projector on V for the operator $L_{s,z}$. As before, denote $L_{s,z}P_{s,z} = B_{s,z}$. Because we have a diagonal block, $B_{s_n,0} = P_{s_n,0}$. Using the previous notations, for some i, the operator $B_{s_n,0}$ is one of the $B^{(i)}_{s_n,0}$s; it has 1 as an eigenvalue with multiplicity $m_i = m$.

In the tensor product $\mathcal{H}^{\wedge m}$, the projector $P^{\wedge m}_{s_n,0}$ is of rank 1. Let $\mathcal{V}_{s_n,0} \in \mathcal{H}^{\wedge m}$ be the associated eigenvector; it also belongs to $Ker(I - L_{s_n,0})^{\wedge m}$. By perturbation theory, we can find a family $\mathcal{V}_{s,z}$, depending analytically on (s, z) in a neighbourhood of $(s_n, 0)$, such that $P^{\wedge m}_{s,z}\mathcal{V}_{s,z} = \mathcal{V}_{s,z}$.

We have

$$(I - L_{s,z})^{\wedge m}\mathcal{V}_{s,z} = \lambda_{s,z}\mathcal{V}_{s,z} \tag{7.14}$$

with $\lambda_{s,z} = det(I - B_{s,z})$.

Similarly there is a family $\mathcal{T}_{s,z}$ in the dual $\mathcal{H}^{*\wedge m}$, depending analytically on the parameters, such that

$$(I - L^*_{s,z})^{\wedge m}\mathcal{T}_{s,z} = \lambda_{s,z}\mathcal{T}_{s,z} \tag{7.15}$$

Differentiating (7.14) once with respect to the parameters, and applying $\mathcal{T}_{s_n,0}$, we get

$$\begin{aligned} &\partial\lambda_{s_n,0}\langle \mathcal{V}_{s_n,0}, \mathcal{T}_{s_n,0}\rangle + \lambda_{s_n,0}\langle \partial\mathcal{V}_{s_n,0}, \mathcal{T}_{s_n,0}\rangle = \\ &\qquad \langle (\partial(I - L_{s_n,0})^{\wedge m})\mathcal{V}_{s_n,0}, \mathcal{T}_{s_n,0}\rangle + \langle (I - L_{s_n,0})^{\wedge m})\partial\mathcal{V}_{s_n,0}, \mathcal{T}_{s_n,0}\rangle \end{aligned} \tag{7.16}$$

Because $(I - L^*_{s_n,0})^{\wedge m}\mathcal{T}_{s_n,0} = \lambda_{s_n,0}\mathcal{T}_{s_n,0}$, the second term on each side of (7.16) are equal, and (7.16) amounts to

$$\partial\lambda_{s_n,0}\langle \mathcal{V}_{s_n,0}, \mathcal{T}_{s_n,0}\rangle = \langle (\partial(I - L_{s_n,0})^{\wedge m})\mathcal{V}_{s_n,0}, \mathcal{T}_{s_n,0}\rangle \tag{7.17}$$

This last term vanishes if $m > 1$, and thus we see that $\partial\lambda_{s_n,0} = 0$. Iterating this procedure, we see that any derivative of order $< m$ of $\lambda_{s,z}$ vanishes at $(s_n, 0)$.

This proves, in particular, that the singularity of the function (7.13) at s_n is at most a pole, and that the residue we are interested in is

$$m \frac{\partial_z \partial_s^{m-1}\lambda_{s_n,0}}{\partial_s^m \lambda_{s_n,0}}$$

as announced earlier.

Iteration of this procedure (differentiate (7.14), then apply $\mathcal{T}_{s_n,0}$) gives after m steps:

$$(\partial_z \partial_s^{m-1} \lambda_{s_n,0}) \langle \mathcal{V}_{s_n,0}, \mathcal{T}_{s_n,0} \rangle = \langle (\partial_z \partial_s^{m-1} (I - L_{s_n,0})^{\wedge m}) \mathcal{V}_{s_n,0}, \mathcal{T}_{s_n,0} \rangle$$
$$= (-1)^m (m-1)! \sum_{k=0}^{m-1} \langle (\partial_s L)^{\wedge k} \wedge \partial_z L \wedge (\partial_s L)^{\wedge m-1-k} \mathcal{V}_{s_n,0}, \mathcal{T}_{s_n,0} \rangle \quad (7.18)$$

The terms where L is not been differentiated disappear, because $1 - L^*_{s_n,0}$ vanishes on $\mathcal{T}_{s_n,0}$. Similarly,

$$\partial_s^m \lambda_{s_n,0} = (-1)^m m! \langle (\partial_s L)^{\wedge m} \mathcal{V}_{s_n,0}, \mathcal{T}_{s_n,0} \rangle$$

We note that $\partial_z L = L \circ M_A$ (where M_A denotes multiplication by A) and $\partial_s L = L \circ M_\tau$. Remembering that $L^*_{s_n,0} \mathcal{T}_{s_n,0} = \mathcal{T}_{s_n,0}$ we can rewrite the last two expressions as

$$\partial_z \partial_s^{m-1} \lambda_{s_n,0} = (-1)^m (m-1)! \sum_{k=0}^{m-1} \langle M_\tau^{\wedge k} \wedge M_A \wedge M_\tau^{\wedge m-1-k} \mathcal{V}_{s_n,0}, \mathcal{T}_{s_n,0} \rangle$$

and

$$\partial_s^m \lambda_{s_n,0} = (-1)^m m! \langle M_\tau^{\wedge m} \mathcal{V}_{s_n,0}, \mathcal{T}_{s_n,0} \rangle.$$

Now, we can choose to write $\mathcal{T}_{s_n,0}$ as

$$\mathcal{T}_{s_n,0} = \wedge_{r_j = r_n} \chi_{(l)}^{ir_j}$$

and $\mathcal{V}_{s_n,0}$ as

$$\mathcal{V}_{s_n,0} = \wedge_{r_j = r_n} \chi_{(r)}^{ir_j},$$

where $\chi_{(l)}^{ir_j}, \chi_{(r)}^{ir_j}$ are associated to T_{ir_j} by the formulae (7.11), (7.12).

For $r_j = r_k = r_n$, we have

$$\langle \tau \chi_{(l)}^{ir_j}, \chi_{(r)}^{ir_k} \rangle = \int \tau(z_2) \frac{T_{ir_j}(dz_1) T_{ir_k}(dz_2)}{|z_1 - z_2|^{2s_n}}$$
$$= \int (\mathcal{R}\mathbb{1})(z_2) \frac{T_{ir_j}(dz_1) T_{ir_k}(dz_2)}{|z_1 - z_2|^{2s_n}} = \mu_0(s_n)^{-1} \langle \phi_{ir_j}, \phi_{ir_k} \rangle,$$

by the formulae of Part 5 (which could as well be applied for two different eigenfunctions of the same eigenvalue). Because the basis (ϕ_{ir_j}) is orthonormal, this coefficient vanishes except for $j = k$.

Similarly,

$$\langle A \chi_{(l)}^{ir_j}, \chi_{(r)}^{ir_k} \rangle = \int A(z_1, z_2) \frac{T^{ir_j}(dz_1) T_{ir_k}(dz_2)}{|z_1 - z_2|^{2s_n}}$$
$$= \int (\mathcal{R}a)(z_1, z_2) \frac{T_{ir_j}(dz_1) T_{ir_k}(dz_2)}{|z_1 - z_2|^{2s_n}},$$

and if $j = k$ this is exactly the Patterson-Sullivan distribution applied to a.

We finally find the expression of the residue.

$$m\frac{\partial_z\partial_s^{m-1}\lambda_{s_n,0}}{\partial_s^m\lambda_{s_n,0}} = \sum_{r_j=r_n}\frac{\int(\mathcal{R}a)(z_1,z_2)\frac{T_{ir_j}(dz_1)T_{ir_j}(dz_2)}{|z_1-z_2|^{2s_n}}}{\int(\mathcal{R}1)(z_1,z_2)\frac{T_{ir_j}(dz_1)T_{ir_j}(dz_2)}{|z_1-z_2|^{2s_n}}}$$

which is what we expected in terms of Patteron-Sullivan distributions.

□

8. Classical Selberg trace formalism

We now begin the Selberg trace formalism proof of Theorem 1.3 (ii). To prepare for the proof, we review the standard theory of the Selberg zeta function and trace formula and then give a non-standard proof which will be generalized in the next section.

As above, we denote by $\{\phi_{ir_k}\}$ an orthonormal basis of $\triangle$-eigenfunctions on $\Gamma\backslash G/K$, with associated eigenvalues $\lambda_k = s_k(1-s_k)$ with $s_k = \frac{1}{2} + ir_k$. In particular the trivial eigenvalue $\lambda_0 = 0$ corresponds to $s_0 = 0, 1$ and $r_0 = \pm\frac{i}{2}$.

8.1. Standard Selberg zeta function

We now review the analytic continuation and polar analysis of the Selberg zeta function. We refer to [V] for background.

The Selberg zeta function is defined by

$$Z(s) = \Pi_{\{P\}}\Pi_{k=0}^{\infty}(1 - N(P)^{-s-k}), \ \ \Re e\ s > 1$$

where $\{P\}$ runs over conjugacy classes of primitive hyperbolic elements and where $N(P) = e^{L_P}$ where L_P is the length of the corresponding geodesic.

The logarithmic derivative of the Selberg zeta function $\frac{1}{s-1/2}\frac{Z'}{Z}(s)$ is defined for $\Re e\ s > 1$ by the formula (see [V], (5.1.5))

$$\mathcal{Z}(s;1) := \frac{1}{s-1/2}\frac{Z'}{Z}(s) = \sum_{\gamma}\frac{L_{\gamma_0}}{\sinh L_\gamma/2}e^{-(s-1/2)L_\gamma}.$$

In this formula, we sum over all closed orbits γ of the geodesic flow and L_γ is the (positive) length of γ.

Theorem 8.1. $\frac{Z'(s)}{Z(s)}$ *admits a meromorphic continuation to* $\mathbf{C}$ *with poles at the points* $s = s_n$ *together with the 'trivial poles' at* $s = -k$, $k = 0, 1, 2, 3, \ldots$.

Proof. We review a few features of the standard proof to draw attention to some important technical issues which might be confusing for the more general versions to come. By definition, we have for $\Re e\ s > 1$,

$$\frac{1}{s-1/2}\frac{Z'(s)}{Z(s)} - \frac{1}{a-1/2}\frac{Z'(a)}{Z(a)} = \sum_\gamma \frac{L_{\gamma_0}}{\sinh L_\gamma/2}\Big\{\frac{1}{2s-1}e^{-(s-1/2)L_\gamma} - \frac{1}{2a-1}e^{-(a-1/2)L_\gamma}\Big\}.$$

See for instance [V], (5.1.5).

To analytically continue the formula, one applies the Selberg trace formula (cf. [V] Theorem 5.5.1) with the test function

$$h(\frac{1}{4}+r^2;s;a)=\frac{1}{(s-\frac{1}{2})^2+r^2}-\frac{1}{(a-\frac{1}{2})^2+r^2}.$$

The Fourier transform of $h(\frac{1}{4}+r^2;s;a)$ is

$$g(u;s;a)=\frac{1}{2s-1}e^{-(s-\frac{1}{2})|u|}-\frac{1}{2a-1}e^{-(a-\frac{1}{2})|u|}.$$

We note that the rate of decay of $h(\frac{1}{4}+r^2;s;a)$ as $r\to\infty$ reflects the singularity of $|u|$ at $u=0$. In the case of a smooth compact quotient, the result is (see [V] Theorem 5.1.1; see also [Sa1])

$$\begin{aligned}\frac{1}{(s-\frac{1}{2})}\frac{Z'(s)}{Z(s)}-\frac{1}{(a-\frac{1}{2})}\frac{Z'(a)}{Z(a)} &= \frac{Vol(\Gamma\backslash G)}{\pi}\sum_{k=0}^{\infty}\left(\frac{1}{s+k}-\frac{1}{a+k}\right)\\ &+\sum_{n=0}^{\infty}\left(\frac{1}{(s-\frac{1}{2})^2+r_n^2}-\frac{1}{(a-\frac{1}{2})^2+r_n^2}\right).\end{aligned} \tag{8.1}$$

We note that the eigenvalue series on the right side would diverge if we only used the formula for $\frac{Z'(s)}{Z(s)}$, but it converges (away from poles) if we subtract $\frac{Z'(a)}{Z(a)}$ or take one derivative.

These formulae give a meromorphic continuation of $\frac{Z'(s)}{Z(s)}$ to $\mathbf{C}$ and show that the poles occur at values of s for which there exists an eigenvalue λ_n satisfying $\lambda_n=s(1-s)$, or at negative integers.

8.2. Convolution operator approach

As sketched above, the Selberg trace formula involves a Fourier transform duality. We will need a more group theoretic approach for the generalizations in the next section, namely the approach in [GGP] to the Selberg trace formula as a formula for the trace of the convolution operator corresponding to a K-bi-invariant function χ.

We denote by $S_{0,0}(G)$ the continuous functions satisfying $\chi(k_1gk_2)=\phi(g)$ for all $k_1,k_2\in K$. The associated convolution operator is defined by

$$R_\chi=\int_G\chi(g)R_g dg,$$

where $R_gf(x)=f(xg)$. There exists a unique (up to scalars) eigenfunction Ψ_s of Ω of eigenvalue $s(1-s)$ in $S_{0,0}$. The spherical transform $S:C_0^\infty(G)\cap S_{0,0}\to PW_m$ is defined by

$$Sf(s)=\int_G f(g)\Phi_s(g)dg.$$

Its range is the subspace of the Paley-Wiener space

$$PW(\mathbf{C})=\{f\in\mathcal{O}(\mathbf{C}):\exists k\;\forall N>0:|f(x+iy)|\le Ce^{k|x|}(1+|y|)^{-N}\} \tag{8.2}$$

with a certain symmetry which we will not need to recall here (see [Z], p.31). Here, $\mathcal{O}(\mathbf{C})$ denotes the holomorphic functions on $\mathbf{C}$.

We also denote the Mellin transform $M : C_0^\infty(A) \to PW(\mathbf{C})$ by

$$Mf(s) = \int_0^\infty f(a) a^s \frac{da}{a},$$

where we identify $f(a)$ as a function of the top diagonal entry of a. Note the non-standard sign of the exponent, which is chosen to be consistent with [L, Z].

The basic Selberg trace formula for a smooth compact quotient (in the form stated in [GGP]) states that

$$\sum_{r_k} S\chi(2ir_k) = Vol(\Gamma\backslash G)\chi(e) + \sum_{\{\gamma\}} \int_{G_\gamma\backslash G} \chi(g^{-1}\gamma g)dg, \quad \chi \in S_{0,0} \tag{8.3}$$

where the sum runs over the principal and complementary series representations (counted with multiplicity), where G_γ is the centralizer of γ in G (similarly for Γ) For $\gamma \neq e$, $\Gamma_\gamma\backslash G_\gamma$ is a closed geodesic.

The orbital integral on the right side of (8.3) may be expressed in terms of the so-called Harish-Chandra transform as follows: If $\chi \in S_{00}$, there exists χ^D on $D = G/K$ such that $\chi(g) = \chi^D(g \cdot 0)$ where $\chi^D(re^{i\theta}) = \chi^D(r)$. In the proof of [Z], Proposition 2.6, it is shown that $\chi(n_u^{-1} a n_u) = \chi^D(\left|\frac{u+i}{u+i\omega}\right|)$, with $\omega = \frac{a+a^{-1}}{a-a^{-1}}$. With some routine manipulation ([Z], pages 55-56), we get

$$H\chi(a) \quad = \quad |a - a^{-1}||\omega| \int_{\mathbf{R}} \chi^D(\left(1 + \tfrac{v-1}{u^2+1}\right)^{1/2}) du, \tag{8.4}$$

and thus

$$\int_{G_\gamma\backslash G} \chi(g^{-1}\gamma g)dg = \frac{Vol(\Gamma_\gamma\backslash G_\gamma)}{|a_\gamma - a_\gamma^{-1}|} H\chi(a_\gamma), \tag{8.5}$$

where $Vol(\Gamma_\gamma\backslash G_\gamma)$ is the length of the closed geodesic. We further have $S\chi = MH\chi$, so we finally obtain

$$\sum_{r_k} MH\chi(2ir_k) = Vol(\Gamma\backslash G)\chi^D(0) + \sum_{\{\gamma\}} \frac{L_{\gamma_0}}{\sinh L_\gamma/2} H\chi(a_\gamma). \tag{8.6}$$

This approach leads most naturally to the zeta function

$$\mathcal{R}(s; 1) := \sum_\gamma \left(\frac{L_{\gamma_0}}{\sinh L_{\gamma/2}} \right) (\cosh L_\gamma/2)^{-2(s-1/2)}. \tag{8.7}$$

In generalizations to non-constant automorphic forms, we begin with (8.7) and then relate it (and its generalizations (9.5) to non-constant automorphic forms) to the usual zeta functions $\mathcal{Z}(s; \sigma)$.

9. Dynamical zeta functions and Selberg trace formalism

This section is concerned with the zeta functions

$$\mathcal{Z}(s; \sigma) := \sum_\gamma \left(\frac{\int_{\gamma_0} \sigma ds}{\sinh L_{\gamma/2}} \right) e^{-(s-1/2)|L_\gamma|}, \quad (\Re e\ s > 1). \tag{9.1}$$

Theorem 9.1. *For each automorphic form* $\sigma = \phi_{ir_k}, X_+\phi_{ir_k}, \psi_m$, $\mathcal{Z}(s;\sigma)$ *is absolutely convergent in* $\Re e\ s > 1$ *and admits a meromorphic continuation to* $\mathbf{C}$. *Except for the trivial representation* $\sigma \equiv 1$, *the only poles in* $\Re e\ s > 0$ *occur at values* $s = \frac{1}{2} + ir$ *for which* $\frac{1}{4} + r^2$ *is an eigenvalue of* $\triangle$, *and the residue is given by*

$$Res_{s=\frac{1}{2}+ir}\ \mathcal{Z}(s;\sigma) = \mu_0(\frac{1}{2} + ir) \sum_{j:\ r_j^2=r^2} \int_{\Gamma\backslash G} \sigma dPS_{ir_j}.$$

This proves a special case of Theorem 1.3 in which the function a has components in a finite number of irreducible representations. We briefly sketch the extension to analytic symbols in the final section.

The proofs are based on a generalized Selberg trace formula introduced in [Z] for the traces $Tr\sigma R_\chi$ on $L^2(\Gamma\backslash G)$ of the composition of R_χ with multiplication by σ. Here, σ is a Casimir eigenfunction of weight m and R_χ is a convolution operator with kernel $\chi \in S_{m,n}(G)$, where $S_{m,n}$ denotes the functions $\chi(g)$ on G satisfying $\chi(k_\psi g k_\theta) == e^{im\psi}e^{in\theta}\chi(g)$, where $k_\theta = exp\theta W \in K$. The eigenspaces of Ω on $S_{m,n}(G)$ are one-dimensional, spanned by the *spherical function* $\Phi_{m,n,s}$ of Ω-eigenvalue $s(1-s)$. We will only be considering the case $n = 0$, and denote the associated normalized spherical function by $\Phi_{m,s}$. Our normalization follows [H, Z]. The spherical transform $S_m : C_0^\infty(G) \cap S_{m,0} \to PW_m$ is defined by

$$S_m f(s) = \int_G f(g)\Phi_{-m,s}(g)dg.$$

Its range is the subspace of the Paley-Wiener space (8.2) with a symmetry depending on m which we will not need to recall here (see [Z], p.31).

We will also need a variety of Harish-Chandra transforms which depend on the weight m and also on the type of representation $\mathcal{P}_{ir}, \mathcal{D}_m^+$. There is a canonical one, defined as follows: Let $\chi \in S_{m,0}$ and let (see [Z] page 57 for (i) and page 49 for (ii)):

$$H_m\chi(a) \quad = \quad |a - a^{-1}| \int_{-\infty}^{\infty} \chi^D(\tfrac{u+i}{u+i\omega})du, \quad (\omega = \tfrac{a+a^{-1}}{a-a^{-1}}). \tag{9.2}$$

Here, if $\chi \in S_{m,0}$ then there exists χ^D on $D = G/K$ such that $\chi(g) = \chi^D(g \cdot 0)$ where $\chi^D(re^{i\theta}) = e^{i\frac{m}{2}\theta}\chi_D(r)$. In the proof of [Z], Proposition 2.6, it is shown that

$$\chi(n_u^{-1}an_u) = \chi^D(\tfrac{u+i}{u+i\omega}) = e^{i\frac{m}{2}\theta(a,u)}\chi^D(\left|\tfrac{u+i}{u+i\omega}\right|)$$

with

$$e^{i\theta(a,u)} = \frac{(u+i)(u^2+\omega^2)^{1/2}}{(u+i\omega)(u^2+1)^{1/2}}, \text{ where } \omega = \frac{a+a^{-1}}{a-a^{-1}}, \ v = \omega^{-2}. \tag{9.3}$$

With some routine manipulation (see [Z], pages 55-56), we get

$$H_m\chi(v) \quad = \quad |a-a^{-1}||\omega|v^{m/4} \int_{\mathbf{R}}(1 + \tfrac{v-1}{u^2+1})^{-m/4}\chi^D(\left(1 + \tfrac{v-1}{u^2+1}\right)^{1/2})du. \tag{9.4}$$

From the Selberg trace formalism viewpoint, it turns out to be most natural to work first with auxiliary dynamical zeta-functions $\mathcal{R}(s;\sigma)$ that do not seem to arise in the thermodynamic formalism. When σ_m has weight m we put

$$\mathcal{R}(s;\sigma_m) := \sum_{\gamma} \left(\frac{\int_{\gamma_0} \sigma_m}{\sinh L_{\gamma/2}} \right) (\tanh L_\gamma/2)^{m/2} (\cosh L_\gamma/2)^{-2(s-1/2)}. \tag{9.5}$$

We then express $\mathcal{Z}(s;\sigma)$ in terms of $\mathcal{R}(s;\sigma)$ to obtain results on the analytic continuation of the latter. This somewhat circuitous route comes about because the trace formula is on the 'quantum level' and therefore does not quite produce the 'classical' zeta-function.

9.1. Forms of weight 0 in $\mathcal{P}_{ir}$

In this section, we prove Theorem 1.3 for the case $\sigma = \phi_{ir_k}$.

In this case the auxiliary zeta function has the form

$$\mathcal{R}(s;\phi_{ir_k}) := \sum_{\gamma} \left(\frac{\int_{\gamma_0} \phi_{ir_k}}{\sinh L_{\gamma/2}} \right) (\cosh L_\gamma/2)^{-2(s-1/2)}. \tag{9.6}$$

Theorem 9.2. *$\mathcal{R}(s;\phi_{ir_k})$ admits a meromorphic continuation to $\mathbf{C}$ with poles at $s = \frac{1}{2} + ir - k, k = 0,1,2,\dots,$ where $\frac{1}{4} + r^2$ is an eigenvalue of $\triangle$, and with*

$$Res_{s=\frac{1}{2}+ir}\, \mathcal{R}(s;\phi_{ir_k}) = \mu_0(ir + \frac{1}{2}) \sum_{j:\, r_j^2 = r^2} \frac{1}{2} \langle \phi_{ir_k}, PS_{r_j} \rangle \,.$$

Proof. We assume throughout that $\phi_{ir_k} \perp 1$, so that the identity term on the $\sum_{\gamma\in\Gamma}$ side of the trace formula vanishes and so that the trivial representation term with $r = \frac{i}{2}$ also vanishes. After the proof, we remark on the case $\phi_{ir_k} \equiv 1$.

By Proposition 2.12 of [Z] (applied in the continuous series case), we have

$$\sum_{n=0}^{\infty} \langle Op(\phi_{ir_k})\phi_{ir_n}, \phi_{ir_n} \rangle M H_0 \chi(2ir_n) = \sum_{\gamma} \left(\frac{\int_{\gamma_0} \phi_{ir_k}}{\sinh L_{\gamma/2}} \right) H^c_{ir_k} \chi(a_\gamma), \tag{9.7}$$

where H_0 is defined by (9.2)-(9.4), and where (see [Z] page 57 for (i) and page 49 for (ii)):

$$H^c_{ir_k}\chi(a) \quad := \quad |a - a^{-1}| \int_{-\infty}^{\infty} F_{ir_k,0}(\tfrac{u-i}{-2i}) \chi^D(\tfrac{u+i}{u+i\omega}) du. \tag{9.8}$$

Here, $a = e^{L/2}$ and $F_{ir_k,0}(\frac{u-i}{-2i})$ is defined in (5.7). We note that the identity term on the right side vanishes by orthogonality.

Remark 9.1. (i) We note that we do not use Proposition 2.10 of [Z], which gives a less convenient zeta function. Although Proposition 2.12 of [Z] is only stated for symbols in the discrete series, it is valid for the continuous series as long as we use the corresponding expressions (given in [Z] Corollary 2.4) for the integrals $I_\gamma(\sigma)(n_u)$ in [Z](2.2).

(ii) A priori, the right side of (9.8) should also contain the term

$$|a-a^{-1}|\int_{-\infty}^{\infty} G_{ir_k,0}(\frac{u-i}{-2i})\chi^D(\frac{u+i}{u+i\omega})du,$$

but χ^D is a radial function since it has weight zero and $\chi^D(\left|\frac{u+i}{u+i\omega}\right|)$ is even in u while $G_{ir_k,0}(\frac{u-i}{-2i})$ is odd. Hence this integral vanishes (cf. Proposition 2.7 of [Z]).

By (9.4), we have

$$\begin{cases} H_0\chi(v) = |a-a^{-1}||\omega| \int_{\mathbf{R}} \chi^D(\left(1+\frac{v-1}{u^2+1}\right)^{1/2})du, \\ \\ H^c_{2ir_k}\chi(v) = |a-a^{-1}||\omega| \int_{\mathbf{R}} F_{ir_k,0}(\frac{u-i}{-2i})\chi^D(\left(1+\frac{v-1}{u^2+1}\right)^{1/2})du. \end{cases} \tag{9.9}$$

We now define $\chi_s(g) \in S_{0,0}$ by the rule that

$$\chi_s^D(r) := \frac{(1-r^2)^s}{\mu^c_{ir_k}(s)}, \quad 0 \le r \le 1.$$

Using (9.9) and the fact that $|a-a^{-1}||\omega| = (a+a^{-1})$, we obtain

$$\begin{cases} (i)\ H^c_{2ir_k}\chi_s(a) = (a+a^{-1})^{-2(s-1/2)}, \\ \\ (ii)\ H_0\chi_s(a) = \frac{\mu_0(s)}{\mu_{ir_k}(s)}\ \left(a+a^{-1}\right)^{-2(s-1/2)} \end{cases}. \tag{9.10}$$

If we substitute χ_s into the right side of the trace formula (9.7), we obtain the desired zeta-function $\mathcal{R}(s;\phi_{ir_k})$. Therefore, the left side of the trace formula (9.7) gives a meromorphic continuation of $\mathcal{R}(s;\phi_{ir_k})$. By Theorem 1.3, we have

$$\langle Op(\phi_{ir_k})\phi_{ir}, \phi_{ir}\rangle = \langle \phi_{ir_k}, PS_{ir}\rangle \mu^c_{ir_k}(\tfrac{1}{2}+ir),$$

hence

$$\mathcal{R}(s;\phi_{ir_k}) = \sum_{n=0}^{\infty} \langle \phi_{ir_k}, PS_{ir_n}\rangle \mu^c_{ir_k}(\frac{1}{2}+ir_n)\ MH_0\chi_s(2ir_n). \tag{9.11}$$

By (9.10(ii)) we have

$$\begin{aligned} MH_0\chi_s(2ir) &= \frac{\mu_0(s)}{\mu^c_{ir_k}(s)} \int_0^{\infty} a^{2ir}(a+a^{-1})^{-2(s-1/2)}\frac{da}{a} \\ &= \frac{\mu_0(s)}{\mu^c_{ir_k}(s)} \int_{-\infty}^{\infty} e^{2irt}(\cosh t)^{-2(s-1/2)}dt \\ &= \frac{\mu_0(s)}{\mu^c_{ir_k}(s)} \frac{\Gamma(s-(\frac{1}{2}+ir))\Gamma(s-(\frac{1}{2}-ir))}{\Gamma(2s-1)}. \end{aligned}$$

For the last line we refer to [Z] (p. 60).

In conclusion, we obtain (at least formally)

$$\mathcal{R}(s;\phi_{ir_k}) = \sum_{n=0}^{\infty} \langle \phi_{ir_k}, PS_{ir_n}\rangle \frac{\mu_0(s)\mu^c_{ir_k}(\frac{1}{2}+ir_n)}{\mu^c_{ir_k}(s)} \frac{\Gamma(s-(\frac{1}{2}+ir_n))\Gamma(s-(\frac{1}{2}-ir_n))}{\Gamma(2s-1)} \tag{9.12}$$

We note that $\frac{\Gamma(s-(\frac{1}{2}+ir))\Gamma(s-(\frac{1}{2}-ir))}{\Gamma(2s-1)} = B(s-(\frac{1}{2}+ir), s-(\frac{1}{2}-ir))$. As above, we assume that $\phi_{ir_k} \perp 1$, so that the trivial representation term vanishes. Regarding the convergence of the right side, we note that by (5.6) and (2.10), as $|r_n| \to \infty$,

$$\begin{cases} \mu^c_{ir_k}(\frac{1}{2}+ir_n) \sim r_n^{-1/2} \\ \Gamma(s-(\frac{1}{2}+ir_n))\Gamma(s-(\frac{1}{2}-ir_n)) \sim e^{-\frac{\pi}{2}(|\Im s + r_n| + |\Im s - r_n|)} \\ \qquad\qquad \times |r_n + \Im s|^{-\Re e(s)-1}|r_n - \Im s|^{-\Re e(s)-1}. \end{cases}$$

Since $\langle \phi_{ir_k}, PS_{ir_n}\rangle = O_{r_k}(r_n^{\frac{1}{2}})$ as $n \to \infty$ (or equivalently, $\langle \phi_{ir_k}, \widehat{PS}_{ir_n}\rangle = O_{r_k}(1)$), it follows that the series converges absolutely in the critical strip away from the poles and defines a meromorphic function.

There are simple poles at $s = \frac{1}{2} \pm ir_n$ where $\frac{1}{4} + r_n^2$ is an eigenvalue of $\triangle$. In the case where the multiplicity of the eigenvalue equals one, the residue at $s = \frac{1}{2} + ir_n$ equals

$$\begin{aligned} \langle \phi_{ir_k}, PS_{ir_n}\rangle \frac{\mu_0(\frac{1}{2}+ir_n)\mu^c_{ir_k}(\frac{1}{2}+ir_n)}{\mu^c_{ir_k}(\frac{1}{2}+ir_n)} \frac{\Gamma(2ir_n)}{\Gamma(2ir_n)} &= \mu_0(\tfrac{1}{2}+ir_n)\langle \phi_{ir_k}, PS_{ir_n}\rangle \\ &= \langle \phi_{ir_k}, \widehat{PS}_{ir_n}\rangle, \end{aligned}$$

as stated. In the case of a multiple eigenvalue one sums over an orthonormal basis of the eigenspace.

□

9.2. $\mathcal{Z}(s;\phi_{ir_k})$

Now we deduce properties of $\mathcal{Z}(s;\phi_{ir_k})$ from those of $\mathcal{R}(s;\phi_{ir_k})$.

We introduce the measure

$$d\Theta(L;\phi_{ir_k}) = \sum_\gamma \left(\frac{\int_{\gamma_0} \phi_{ir_k} ds}{\sinh L_\gamma/2} \right) \delta(L - L_\gamma).$$

We note that

$$\begin{cases} \mathcal{Z}(s;\phi_{ir_k}) = \int_0^\infty e^{-(s-\frac{1}{2})L} \ d\Theta(L;\phi_{ir_k}), \\ \\ \mathcal{R}(s;\phi_{ir_k}) = \int_0^\infty \left((\frac{e^{L/2}+e^{-L/2}}{2})^2 \right)^{-(s-\frac{1}{2})} d\Theta(L;\phi_{ir_k}). \end{cases} \tag{9.13}$$

Lemma 9.3. *We have:*

$$\mathcal{Z}(s;\phi_{ir_k}) = \sum_{n=0}^{\infty} B_m(s,n)\mathcal{R}(s+n;\phi_{ir_k}),$$

where

$$B_m(s,n) = 2^{-s+\frac{1}{2}} 2^n \left\{ \sum_{m,k_1,\dots,k_m=0;k_1+\cdots k_m=n}^{\infty} \binom{2s-1}{m} \binom{\frac{1}{2}}{k_1+1} \cdots \binom{\frac{1}{2}}{k_m+1} \right\}$$

Proof. By elementary manipulation, we have

$$\mathcal{Z}(s, \phi_{ir_k}) = \int_0^\infty \left(1+e^{-L}\right)^{2s-1} d\Theta(L; s; \phi_{ir_k}), \tag{9.14}$$

where

$$d\Theta(L, s; \phi_{ir_k}) = \sum_\gamma \left(\frac{\int_{\gamma_0} \phi_{ir_k}}{\sinh L_\gamma/2} \right) \left((\cosh L_\gamma/2)^2\right)^{-(s-\frac{1}{2})} \; \delta(L - L_\gamma).$$

We then change variables to $y = (\cosh L/2)^2$, and note that $e^{-L/2} = \sqrt{y} - \sqrt{y-1}$ to obtain,

$$\begin{aligned} \mathcal{Z}(s, \phi_{ir_k}) &= \int_0^\infty \left(1 + (\sqrt{y} - \sqrt{y-1})^2\right)^{2s-1} d\Psi(y; s; \phi_{ir_k}) \\ &= \int_0^\infty (2y)^{2s-1} \left(1 - \sqrt{1-\tfrac{1}{y}}\right)^{2s-1} d\Psi(y; s; \phi_{ir_k}) \end{aligned}, \tag{9.15}$$

where

$$d\Psi(y; s; \phi_{ir_k}) = \sum_\gamma \left(\frac{\int_{\gamma_0} \phi_{ir_k}}{\sinh L_\gamma/2} \right) (y_\gamma)^{-(s-\frac{1}{2})} \; \delta(y - y_\gamma).$$

By repeated use of the binomial theorem, there exist coefficients $B_m(s,n)$ such that

$$\left(1 - \sqrt{1 - \frac{1}{y}}\right)^{2s-1} = y^{-(2s-1)} \sum_{n=0}^\infty B_m(s,n) y^{-n}.$$

Canceling the factors of $y^{\pm(2s-1)}$, we thus have

$$\begin{aligned} \mathcal{Z}(s, \phi_{ir_k}) &= \sum_{n=0}^\infty B_m(s,n) \int_0^\infty y^{-n} d\Psi(y; s; \phi_{ir_k}) \\ &= \sum_{n=0}^\infty B(s,n) \mathcal{R}(s+n; \phi_{ir_k}). \end{aligned} \tag{9.16}$$

Since the poles of $\mathcal{R}(s+n, \phi_{ir_k})$ are the shifts by $-n$ of the poles of $\mathcal{R}(s, \phi_{ir_k})$, and since the non-trivial poles of $\mathcal{R}(s, \psi_m)$ in $\Re e \; s > 0$ lie only at the points $s = \frac{1}{2} + ir$, only the term $n = 0$ in the series contributes non-trivial poles to the critical strip.

Writing out $B_m(s,0)$ as a sum $\sum_{m,k_1,\dots,k_m=0;k_1+\cdots k_m=n}$, we see that the only term has $m = 0 = k_j$ (for all $j = 1, \dots, m$). Thus,

$$Res_{s=\frac{1}{2}+ir} \mathcal{Z}(s; \phi_{ir_k}) = Res_{s=\frac{1}{2}+ir} \mathcal{R}(s; \phi_{ir_k}).$$

This completes the proof of Theorem 9.1 in the case $\sigma = \phi_{ir_k}$.

□

□

Remark 9.2. As a check on (9.12), we observe that in the case $\phi_{ir_k} \equiv 1$, $\langle \phi_{ir_k}, \widehat{PS}_{ir_n} \rangle = 1$ for all n, $\mu_{ir_k}(s) = \mu_0(s)$ and we get

$$\begin{aligned} \mathcal{R}(s;1) &= \frac{Vol(\Gamma\backslash G)}{2\pi} \int_{\mathbf{R}} \frac{\Gamma(s-(\frac{1}{2}+ir))\Gamma(s-(\frac{1}{2}-ir))}{\Gamma(2s-1)} r(\tanh \pi r) dr \\ &\quad + \sum_{n=0}^{\infty} \frac{\Gamma(s-(\frac{1}{2}+ir_n))\Gamma(s-(\frac{1}{2}-ir_n))}{\Gamma(2s-1)}. \end{aligned} \tag{9.17}$$

The series converges rapidly to a meromorphic function with simple poles at $s = \frac{1}{2} \pm r_n - k$ $(k = 0, 1, 2, \dots)$, the residue at $s = \frac{1}{2} \pm r_n - k$ being $\frac{(-1)^k}{k!}$. Thus, Lemma 9.3 shows that $\mathcal{Z}(s;1)$ has simple poles in the critical strip with residues equal to 1. The formula (9.17) also follows from the standard Selberg trace formula (Fourier transform duality, [V] Theorem 4.3.6) by using the integral formula (cf. [WW], Exercise 24)

$$B(s-\frac{1}{2}-ir_n, s-\frac{1}{2}+ir_n) = \frac{1}{4^{s-\frac{1}{2}}} \int_{\mathbf{R}} \frac{\cos(2ir_n u) du}{\cosh^{2s-1}(u)}$$

and the fact noted above that $\frac{\Gamma(s-(\frac{1}{2}+ir))\Gamma(s-(\frac{1}{2}-ir))}{\Gamma(2s-1)} = B(s-(\frac{1}{2}+ir), s-(\frac{1}{2}-ir))$.

9.3. Forms of weight ± 2 in $\mathcal{P}_{ir}$

In this case both sides of the trace formula equal zero due to time reversibility. By Propositions 2.3 and 3.3, each side of the trace formula equals zero, noting that $\left(\frac{\int_{\gamma_0} X_+ \phi_{ir_k} ds}{\sinh L_{\gamma/2}} \right) + \left(\frac{\int_{\gamma_0^{-1}} X_+ \phi_{ir_k} ds}{\sinh L_{\gamma/2}} \right) = 0$.

9.4. Weight m in $\mathcal{D}^{\pm}_m$

We now prove Theorem 9.1 for $\sigma = \psi_m \in \mathcal{D}^+_m$. The anti-holomorphic discrete series case is simply the complex conjugate and is omitted.

The proof is similar to the case $\mathcal{Z}(s;\phi_{ir_k})$ but involves the higher weight analogue zeta-function:

$$\mathcal{R}(s;\psi_m) := \sum_{\gamma} \left(\frac{\int_{\gamma_0} \psi_m ds}{\sinh L_{\gamma/2}} \right) (\tanh L_\gamma/2)^{m/2} (\cosh L_\gamma/2)^{-2(s-1/2)}. \tag{9.18}$$

We begin the proof with an analysis of its meromorphic continuation.

9.4.1. Meromorphic continuation of $\mathcal{R}(s;\psi_m)$.

Theorem 9.4. *$\mathcal{R}(s;\psi_m)$ admits a meromorphic continuation to* $\mathbf{C}$. *In the critical strip, its poles occur at $s = \frac{1}{2} + ir$ such that $\frac{1}{4} + r^2$ is an eigenvalue of $\triangle$, with residue $\mu_0(\frac{1}{2}+ir) \sum_{j:\, r_j^2 = r^2} \langle \psi_m, PS_{ir_j} \rangle$.*

Proof. We study $\mathcal{R}(s;\psi_m)$ using the trace formula given in [Z], Proposition 2.12:

$$\sum_{n=0}^{\infty} \langle Op(\psi_m)\phi_{ir_n}, \phi_{ir_n} \rangle MH_m\chi(2ir_n) = \sum_{\gamma} \left(\frac{\int_{\gamma_0} \psi_m}{\sinh L_{\gamma/2}} \right) H^d_m \chi(a_\gamma), \tag{9.19}$$

where (see [Z] page 57 for (i) and page 49 for (ii)):

$$\begin{cases} (i)\ H_m\chi(a) &= |a-a^{-1}| \int_{-\infty}^{\infty} \chi^D(\frac{u+i}{u+i\omega})du, \\ (ii)\ H_m^d\chi(a) &= |a-a^{-1}| \int_{-\infty}^{\infty} (u+i)^{-m/2}\chi^D(\frac{u+i}{u+i\omega})du. \end{cases} \tag{9.20}$$

We caution that in the definition of H_m^d (9.20)(ii) we follow a slightly different notation convention in [Z] whereby we multiply the integral by $|a-a^{-1}|$ as for H_m

The integral uses the notation of (9.2)-(9.3). We simplify the expressions in (9.20) by further using these identities to obtain (see also [Z], pages 55-56)

$$\begin{cases} (i)\ H_m\chi(v) = |a-a^{-1}||\omega|v^{m/4} \int_{\mathbf{R}}(1+\frac{v-1}{u^2+1})^{-m/4}\chi^D\left(\left(1+\frac{v-1}{u^2+1}\right)^{1/2}\right)du, \\ (ii)\ H_m^d\chi(v) = |a-a^{-1}||\omega|v^{m/4} \int_{\mathbf{R}}(u+i)^{-m/2}(1+\frac{v-1}{u^2+1})^{-m/4} \\ \qquad\qquad \times\ \chi^D\left(\left(1+\frac{v-1}{u^2+1}\right)^{1/2}\right)du, \end{cases} . \tag{9.21}$$

We now define $\chi_s(g) \in S_{m,0}$ by the rule that

$$r^{-m/2}\chi_s^D(r) := \frac{(1-r^2)^s}{\mu_m^d(s)}, \quad 0 \le r \le 1,$$

where (see [Z], Proposition 3.6)

$$\mu_m^d(s) = \int_{\mathbf{R}}(u+i)^{-m/2}(u^2+1)^{-s}du = \frac{(-i)^{m/2}\pi 2^{2s+2-m/2}\Gamma(-2s+\frac{m}{2})}{-(2s+1-\frac{m}{2})\Gamma(-s)\Gamma(-s+\frac{m}{2})}.$$

Since $(1-v)^s = (a+a^{-1})^{-2s}$, and $|a-a^{-1}||\omega|v^{m/4} = (a+a^{-1})\left(\frac{a-a^{-1}}{a+a^{-1}}\right)^{m/2}$, we have

$$\begin{cases} (i)\ \ H_m^d\chi_s(a) &= \left(\frac{a-a^{-1}}{a+a^{-1}}\right)^{m/2}(a+a^{-1})^{-2(s-1/2)}, \\ (ii)\ \ H_m\chi_s(a) &= \frac{\mu_0(s)}{\mu_m^d(s)}\left(\frac{a-a^{-1}}{a+a^{-1}}\right)^{m/2}(a+a^{-1})^{-2(s-1/2)} \end{cases} . \tag{9.22}$$

It follows first that if we substitute χ_s into the right side of the trace formula (9.19) is the desired zeta-function $\mathcal{R}(s;\psi_m)$. Therefore, the left side of the trace formula (9.19) gives a meromorphic continuation of $\mathcal{R}(s;\psi_m)$. By Theorem 1.3, we have

$$\langle Op(\psi_m)\phi_{ir},\phi_{ir}\rangle \quad = \quad \langle\psi_m, PS_{ir}\rangle\mu_m^d(\tfrac{1}{2}+ir),$$

hence

$$\mathcal{R}(s;\psi_m) = \sum_{n=0}^{\infty}\langle\psi_m, PS_{ir_n}\rangle\mu_m^d(\frac{1}{2}+ir_n)\ MH_m\chi_s(2ir_n). \tag{9.23}$$

By (9.22(ii)) we have

$$\begin{aligned} MH_m\chi_s(2ir) &= \frac{\mu_0(s)}{\mu_m^d(s)}\int_0^\infty a^{2ir}\left(\frac{a-a^{-1}}{a+a^{-1}}\right)^{m/2}(a+a^{-1})^{-2(s-1/2)}\frac{da}{a} \\ &= \frac{\mu_0(s)}{\mu_m^d(s)}\int_{-\infty}^\infty e^{2irt}\,(\tanh t)^{m/2}\,(\cosh t)^{-2(s-1/2)}dt. \end{aligned}$$

Putting things together, we obtain the discrete series analogue of (9.12),

$$\begin{aligned} \mathcal{R}(s;\psi_m) = \sum_{n=0}^\infty \langle \psi_m, PS_{ir_n}\rangle \mu_m^d(\frac{1}{2}+ir_n)\ \frac{\mu_0(s)}{\mu_m^d(s)} \\ \times\int_{-\infty}^\infty e^{2ir_nt}\,(\tanh t)^{m/2}\,(\cosh t)^{-2(s-1/2)}dt. \quad (9.24) \end{aligned}$$

The integral is more complicated than its zero weight analogue, but as $\tanh t = 1+r(t)$ with $r(t)=O(e^{-2|t|})$, we may write

$$\begin{aligned} \int_{-\infty}^\infty e^{2ir_nt}\,(\tanh t)^{m/2}\,(\cosh t)^{-2(s-1/2)}dt = \\ \int_{-\infty}^\infty e^{2ir_nt}(\cosh t)^{-2(s-1/2)}dt + R_2(s,r_n), \quad (9.25) \end{aligned}$$

where

$$R_2(s,r_n) = \int_{-\infty}^\infty e^{2ir_nt}r(t)(\cosh t)^{-2(s-1/2)}dt. \quad (9.26)$$

The first term of (9.25) gives the expression in the weight zero case analyzed above. Hence, the sum over r_n with this term converges, and the poles and residues of $\mathcal{R}(s;\psi_m)$ on $\Re e\ s = 1/2$ due to this term are the same as for

$$\langle \psi_m, PS_{ir_n}\rangle \frac{\mu_0(s)\mu_m^d(\frac{1}{2}+ir_n)}{\mu_m^d(s)}\frac{\Gamma(s-(\frac{1}{2}+ir_n)\Gamma(s-(\frac{1}{2}-ir_n))}{\Gamma(2s-1)}.$$

There are simple poles at $s = \frac{1}{2}+ir_n$ and the residue is $\langle \psi_m, PS_{ir_n}\rangle \frac{\mu_0(\frac{1}{2}+ir_n))\mu_m^d(\frac{1}{2}+ir_n)}{\mu_m^d(\frac{1}{2}+ir_n))} = \langle \psi_m, PS_{ir_n}\rangle \mu_0(\frac{1}{2}+ir_n)$. Summing over an orthonormal basis of lowest weight vectors of $\mathcal{D}_m^+$ gives the stated expression.

To complete the proof, it is only necessary to observe that the second integral $R_2(s,r_n)$ is holomorphic in the region $\Re e(s) > -\frac{1}{2}$. It is also rapidly decaying in r_n. Therefore it does not contribute any poles or residues to $\mathcal{R}(s;\psi_m)$ in the critical strip.

□

9.5. $\mathcal{Z}(s;\psi_m)$

Now we deduce properties of $\mathcal{Z}(s;\psi_m)$ from those of $\mathcal{R}(s;\psi_m)$.

As in the weight zero case, we introduce the measure

$$d\Theta(L;\psi_m) = \sum_\gamma \left(\frac{\int_{\gamma_0}\psi_m ds}{\sinh L_\gamma/2}\right)(\tanh L_\gamma/2)^{m/2}\ \delta(L-L_\gamma).$$

We note that

$$\begin{cases} \mathcal{Z}(s;\psi_m) = \int_0^\infty e^{-(s-\frac{1}{2})L} \ (\tanh L/2)^{-m/2} \ d\Theta(L;\psi_m), \\ \\ \mathcal{R}(s;\psi_m) = \int_0^\infty \left((\frac{e^{L/2}+e^{-L/2}}{2})^2 \right)^{-(s-\frac{1}{2})} d\Theta(L;\psi_m). \end{cases} \tag{9.27}$$

Because the factor $(\tanh L/2)^{-m/2}$ is somewhat inconvenient, we also consider

$$\begin{cases} \tilde{\mathcal{Z}}(s;\psi_m) & = & \int_0^\infty e^{-(s-\frac{1}{2})L} \ d\Theta(L;\psi_m), \\ & = & \sum_\gamma \left(\frac{\int_{\gamma_0} \psi_m ds}{\sinh L_\gamma/2} \right) (\tanh L_\gamma/2)^{m/2} \ e^{-(s-\frac{1}{2})L_\gamma}. \end{cases} \tag{9.28}$$

Lemma 9.5. *We have:*

$$\tilde{\mathcal{Z}}(s;\psi_m) = \sum_{n=0}^\infty B_m(s,n)\mathcal{R}(s+n;\psi_m),$$

where $B_m(s,n)$ is the same as in Lemma 9.3.

Proof. We use similar manipulations as in the weight zero case. We now have

$$\tilde{\mathcal{Z}}(s,\psi_m) = \int_0^\infty \left(1+e^{-L}\right)^{2s-1} d\Theta(L;s;\psi_m), \tag{9.29}$$

where

$$d\Theta(L,s;\psi_m) = \sum_\gamma \left(\frac{\int_{\gamma_0} \psi_m ds}{\sinh L_\gamma/2} \right)$$
$$(\tanh L_\gamma/2)^{m/2} \ \left((\cosh L_\gamma/2)^2\right)^{-(s-\frac{1}{2})} \ \delta(L-L_\gamma).$$

We change variables as before to $y = (\cosh L/2)^2$, and obtain as in (9.15),

$$\tilde{\mathcal{Z}}(s,\psi_m) \quad = \quad \int_0^\infty (2y)^{2s-1} \left(1-\sqrt{1-\frac{1}{y}}\right)^{2s-1} d\Psi(y;s;\psi_m) \ , \tag{9.30}$$

where

$$d\Psi(y;s;\psi_m) = \sum_\gamma \left(\frac{\int_{\gamma_0} \psi_m ds}{\sinh L_\gamma/2} \right) (\tanh L_\gamma/2)^{m/2} \ (y_\gamma)^{-(s-\frac{1}{2})} \ \delta(y-y_\gamma).$$

As in the weight zero case, we then have

$$\begin{aligned} \tilde{\mathcal{Z}}(s,\psi_m) & = \sum_{n=0}^\infty B_m(s,n) \int_0^\infty y^{-n} d\Psi(y;s;\psi_m) \\ & = \sum_{n=0}^\infty B(s,n)\mathcal{R}(s+n;\psi_m). \end{aligned} \tag{9.31}$$

Since the poles of $\mathcal{R}(s+n,\psi_m)$ are the shifts by $-n$ of the poles of $\mathcal{R}(s,\psi_m)$, and since the non-trivial poles of $\mathcal{R}(s,\psi_m)$ in $\Re e \ s > 0$ lie only at the points

$s = \frac{1}{2} + ir$, only the term $n = 0$ in the series contributes non-trivial poles to the critical strip, and as above this term has $m = 0 = k_j$ (for all $j = 1, \dots, m$). Thus,

$$Res_{s=\frac{1}{2}+ir}\tilde{\mathcal{Z}}(s;\psi_m) = Res_{s=\frac{1}{2}+ir}\mathcal{R}(s;\psi_m).$$

□

To complete the proof of the theorem, we now observe that

$$\mathcal{Z}(s,\psi_m) - \tilde{\mathcal{Z}}(s,\psi_m) = \sum_\gamma \left(\frac{\int_{\gamma_0} \psi_m ds}{\sinh L_\gamma/2} \right) \left[1 - (\tanh L_\gamma/2)^{m/2}\right] e^{-(s-\frac{1}{2})L_\gamma}. \tag{9.32}$$

Since $\left[1 - (\tanh L_\gamma/2)^{m/2}\right] = \mathcal{O}(e^{-L_\gamma})$ and since

$$\sum_\gamma \left| \left(\frac{\int_{\gamma_0} \psi_m ds}{\sinh L_\gamma/2} \right) e^{(-s-1+\frac{1}{2})L_\gamma} \right| < \infty, \quad \Re e\ s > 0, \tag{9.33}$$

by the prime geodesic theorem, it follows that $\mathcal{Z}(s,\psi_m)$ has the same poles and residues in the critical strip as $\tilde{\mathcal{Z}}(\psi_m)$.

This completes the proof of Theorem 9.4.

9.6. Meromorphic extension of $\mathcal{Z}$: Proof of Theorem 1.3 for $\mathcal{Z}(s;\sigma)$

By Proposition 2.4, by a similar calculation as in Corollary 6.16, we have

$$\mathcal{Z}(s;\sigma) = \sum_{r_j} \frac{\langle \sigma, \Xi_{ir_j} \rangle}{\langle \phi_{ir_j}, \Xi_{ir_j} \rangle} \mathcal{Z}(s;\phi_{ir_j}) + \sum_{m,\pm} \frac{\langle \sigma, \Xi_m^\pm \rangle}{\langle \psi_m, \Xi_m^\pm \rangle} \mathcal{Z}(s;\psi_m), \quad (\Re e(s) > 1). \tag{9.34}$$

Here, we interchanged the summation over γ and over r_j, which is justified by Proposition 2.5 and the prime geodesic theorem.

Under the assumption that σ has non-trivial projections in only finitely many irreducible represention, the analytic continuation of the sums follows from that of the individual terms, which has been proved in Theorems 9.2 and 9.4.

Remark 9.3. Note that σ may have an infinite number of non-zero Fourier coefficients relative to automorphic (τ, m)- eigenfunctions; it is only in the τ aspect that we assume finiteness.

Remark 9.4. It is natural to ask for the precise conditions on σ, specifically the decay rate of the coefficients

$$\frac{\langle \sigma, \Xi_{ir_j} \rangle}{\langle \phi_{ir_j}, \Xi_{ir_j} \rangle}, \quad \frac{\langle \sigma, \Xi_m^\pm \rangle}{\langle \psi_m, \Xi_m^\pm \rangle}, \tag{9.35}$$

to ensure that $\mathcal{Z}(s;\sigma)$ admits a meromorphic continuation to $\mathbf{C}$. In the introduction, we said that this question is related to estimates on triple products in [BR2, Sa3]. Let us briefly explain the connection.

By Lemmas 9.3 and 9.5 it suffices to prove the meromorphic continuation of the zeta functions

$$\mathcal{R}(s;\sigma) = \sum_{r_k} \frac{\langle \sigma, \Xi_{ir_k}\rangle}{\langle \phi_{ir_k}, \Xi_{ir_k}\rangle} \mathcal{R}(s;\phi_{ir_k}) + \sum_{m,\pm} \frac{\langle \sigma, \Xi_m^{\pm}\rangle}{\langle \psi_m, \Xi_m^{\pm}\rangle} \mathcal{R}(s;\psi_m), \quad (\Re e(s) > 1). \tag{9.36}$$

Since

$$\langle Op(\phi_{ir_k})\phi_{ir_n}, \phi_{ir_n}\rangle = \mu^c_{ir_k}(\frac{1}{2} + ir_n)\, \langle \phi_{ir_k}, PS_{ir_n}\rangle\,,$$

we have

$$\mathcal{R}(s;\phi_{ir_k}) = \sum_{n=0}^{\infty} \langle Op(\phi_{ir_k})\phi_{ir_n}, \phi_{ir_n}\rangle \frac{\mu_0(s)}{\mu^c_{ir_k}(s)} \frac{\Gamma(s-(\frac{1}{2}+ir_n))\Gamma(s-(\frac{1}{2}-ir_n))}{\Gamma(2s-1)}. \tag{9.37}$$

Similarly in the discrete series.

The following is due to Sarnak [Sa3] and (in its stated form) Bernstein-Reznikov [BR2]:

Lemma 9.6. $|\langle Op(\phi_{ir_k})\phi_{ir_n}, \phi_{ir_n}\rangle| \leq C_n e^{-\frac{\pi |r_k|}{2}} \left(\log |r_k|\right)^{3/2}$.

It follows that $C_n | \frac{\langle Op(\phi_{ir_k})\phi_{ir_n}, \phi_{ir_n}\rangle}{\mu^c_{ir_k}(s)} | \leq C_{s,n}(1+|r_k|)^{-2\Re s + \frac{3}{2}}$, where $C_{s,n}$ is uniform on compact sets of $\mathbf{C}$. Thus, the r_k-sum for fixed r_n converges absolutely as long as the coefficients (9.35) decay rapidly enough, and certainly if σ is real analytic. However, there do not seem to exist estimates of the coefficients C_n in Lemma 9.6, and hence no proof that the full (r_k, r_n) sum converges. It seems reasonable at this time that the coefficients C_n could grow to order $e^{\pi r_n}$, which would cancel the Gamma factors and leave the convergence unclear.

Acknowledgments. This work was begun while the first author was visiting Johns Hopkins University as part of the NSF focussed research grant # FRG 0354386. Much of it was written at the Time at Work program of the Institut Henri Poincaré in June, 2005.

References

[Ag] S. Agmon, On the representation theorem for solutions of the Helmholtz equation on the hyperbolic space. *Partial differential equations and related subjects* (Trento, 1990), 1–20, Pitman Res. Notes Math. Ser., 269, Longman Sci. Tech., Harlow, 1992.

[A-P] A. Alvarez-Parrilla, *Explicit geodesic-flow invariant distributions using $SL(2,\mathbf{R})$ ladders*, Int. J. Math. and Math. Sci. **8** (2005), 1299–1315.

[AN] N. Anantharaman, S. Nonnenmacher, *Half-delocalization of the eigenfunctions of the Laplacian,* preprint 2006.

[AZ] N. Anantharaman and S. Zelditch, in progress.

[BR] J. Bernstein and A. Reznikov, *Sobolev norms of automorphic functionals*, Int. Math. Res. Not. **40** (2002) 2155–2174.

[BR2] J. Bernstein and A. Reznikov, Analytic continuation of representations and estimates of automorphic forms. Ann. of Math. (2) 150 (1999), no. 1, 329–352.

[C] F. Chamizo, *Automorphic forms and differentiability properties*, Trans. Amer. Math. Soc. **356** (2004), no. 5, 1909–1935.

[Co] S. Cosentino, *A note on Hölder regularity of invariant distributions for horocycle flows*, Nonlinearity **18** (2005), no. 6, 2715–2726.

[FF] L. Flaminio and G. Forni, *Invariant distributions and time averages for horocycle flows*, Duke Math. J. **119** (2003), no. 3, 465–526.

[GFa] I. M. Gel'fand and S. V. Fomin, *Unitary representations of Lie groups and geodesic flows on surfaces of constant negative curvature* (in Russian), Dokl. Akad. Nauk SSSR 76 (1951), 771–774.

[GF] I.M. Gel' fand, S. V. Fomin, *Geodesic flows on manifolds of constant negative curvature*, Amer. Math. Soc. Transl. **2**, 1 (1955), 49–65.

[GGP] I. M. Gelf́and, M. I. Graev and I. I. Pyatetskii-Shapiro, *Representation theory and automorphic functions*, W. B. Saunders Co., Philadelphia, Pa.-London-Toronto, Ont. 1969.

[GR] I. S. Gradshteyn and I.M. Ryzhik, *Table of integrals, series, and products.* Academic Press, Inc., San Diego, CA, 2000.

[G] V. Guillemin, *Lectures on spectral theory of elliptic operators*, Duke Math. J. **44** (1977), no. 3, 485–517.

[H] S. Helgason, *Topics in harmonic analysis on homogeneous spaces*, Progress in Mathematics, 13. Birkhäuser, Boston, Mass., 1981.

[He] S. Helgason, *Groups and geometric analysis. Integral geometry, invariant differential operators, and spherical functions*, Corrected reprint of the 1984 original. Mathematical Surveys and Monographs, 83. American Mathematical Society, Providence, RI, 2000.

[J] A. Juhl, *Cohomological theory of dynamical zeta functions*, Progress in Mathematics, 194. Birkhäuser Verlag, Basel, 2001.

[K] A. W. Knapp,*Representation theory of semisimple groups. An overview based on examples*, Reprint of the 1986 original. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 2001.

[L] S. Lang, $\mathrm{SL}_2(R)$, Graduate Texts in Mathematics, 105. Springer-Verlag, New York, 1985.

[Lin] E. Lindenstrauss, *Invariant measures and arithmetic quantum unique ergodicity*, (Annals Math., to appear).

[Mar1] B. Marcus, *Ergodic properties of horocycle flows for surfaces of negative curvature*, Ann. of Math. **2**, 105 (1977), 81–105.

[Mar2] B. Marcus, *The horocycle flow is mixing of all degrees*, Invent. Math. **46** (1978), 201–209.

[Mo] T. Morita, *Markov systems and transfer operators associated with cofinite Fuchsian groups*, Ergodic Theory Dynam. Systems **17** (1997), no. 5, 1147–1181.

[McK] H. P. McKean, *Selberg's trace formula as applied to a compact Riemann surface*, Comm. Pure Appl. Math. **25** (1972), 225–246.

[MS] S. D. Miller and W. Schmid, *The highly oscillatory behavior of automorphic distributions for* SL(2), Lett. Math. Phys. **69** (2004), 265–286.

[N] P.J. Nicholls, *The Ergodic Theory of Discrete Groups*, London Math. Soc. Lect. Notes Series 143, Cambridge Univ. Press, Cambridge 143.

[O] J.P. Otal, *Sur les fonctions propres du laplacien du disque hyperbolique*, C. R. Acad. Sci. Paris Sér. I Math. **327** (1998), no. 2, 161–166.

[P] O. S. Parasyuk, *Flows of horocycles on surfaces of constant negative curvature* (in Russian), Uspekhi Mat. Nauk **8**, no. 3 (1953), 125–126.

[Pat0] S. J. Patterson, *The limit set of a Fuchsian group*, Acta Math. **136** (1976), no. 3-4, 241–273.

[Pat1] S. J. Patterson, *Lectures on measures on limit sets of Kleinian groups*, Analytical and geometric aspects of hyperbolic space (Coventry/Durham, 1984), 281–323, London Math. Soc. Lecture Note Ser., 111, Cambridge Univ. Press, Cambridge, 1987.

[Pol] M. Pollicott, *Some applications of thermodynamic formalism to manifolds with constant negative curvature*, Adv. Math. **85** (1991), no. 2, 161–192.

[R] A. Reznikov, *Microlocal lifts of eigenfunctions on hyperbolic surfaces and trilinear invariant functionals*, (preprint, 2004; math.AP/0404294).

[RS] Z. Rudnick and P. Sarnak, *The behaviour of eigenstates of arithmetic hyperbolic manifolds*, Comm. Math. Phys. **161** (1994), no. 1, 195–213.

[Ru87] D. Ruelle, *Resonances for Axiom A flows*, J. Differential Geom. **25** (1987), no. 1, 99–116.

[Rugh92] H. H. Rugh, *The correlation spectrum for hyperbolic analytic maps*, Nonlinearity **5** (1992), no. 6, 1237–1263.

[Rugh96] H. H. Rugh, *Generalized Fredholm determinants and Selberg zeta functions for Axiom A dynamical systems*, Ergodic Theory Dynam. Systems **16** (1996), no. 4, 805–819.

[Sa1] P. Sarnak, *Determinants of Laplacians*, Comm. Math. Phys. **110** (1987), no. 1, 113–120.

[Sa2] P. Sarnak, *Some applications of modular forms*, Cambridge Tracts in Mathematics, 99. Cambridge University Press, Cambridge, 1990.

[Sa3] P. Sarnak, Integrals of products of eigenfunctions, IMRN, no. 6 (1994), 251–260

[Schm] W. Schmid, *Automorphic distributions for* $SL(2,\mathbb{R})$, Conférence Moshé Flato 1999, Vol. I (Dijon), 345–387, Math. Phys. Stud., 21, Kluwer Acad. Publ., Dordrecht, 2000.

[Sh] A. I. Schnirelman, *Ergodic properties of eigenfunctions*, Usp. Mat. Nauk. **29/6** (1974), 181–182.

[Se] C. Series, *The infinite word problem and limit sets in Fuchsian groups*, ETDS **1** (1981), 337–360.

[SV] L. Silberman and A. Venkatesh, *On Quantum unique ergodicity for locally symmetric spaces I*, (math.RT/0407413).

[Su1] D. Sullivan, *On the ergodic theory at infinity of an arbitrary discrete group of hyperbolic motions*, Riemann surfaces and related topics: Proceedings of the 1978

Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978), pp. 465–496, Ann. of Math. Stud., 97, Princeton Univ. Press, Princeton, N.J., 1981.

[Su2] D. Sullivan, *The density at infinity of a discrete group of hyperbolic motions*, Inst. Hautes Études Sci. Publ. Math. **50** (1979), 171–202.

[V] A. B. Venkov, *Spectral theory of automorphic functions*, A translation of Trudy Mat. Inst. Steklov. 153 (1981). Proc. Steklov Inst. Math. 1982, no. 4(153), ix+163 pp. (1983).

[WW] E.T. Whittaker and G. N. Watson, *A course of modern analysis*, Fourth edition. Reprinted Cambridge University Press, New York 1962.

[W] S. A. Wolpert, *Semiclassical limits for the hyperbolic plane*, Duke Math. J. **108** (2001), no. 3, 449–509.

[Z] S. Zelditch, *Trace formula for compact* $\Gamma\backslash \mathrm{PSL}_2(R)$ *and the equidistribution theory of closed geodesics*, Duke Math. J. **59** (1989), no. 1, 27–81.

[Z2] S. Zelditch, *Uniform distribution of eigenfunctions on compact hyperbolic surfaces*, Duke Math. J. **55** (1987), 919–941.

[Z3] S. Zelditch, *Pseudodifferential analysis on hyperbolic surfaces*, J. Funct. Anal. **68** (1986), no. 1, 72–105.

Nalini Anantharaman
École Normale Supérieure de Lyon
46, allée d'Italie
69364 Lyon Cedex 07, France
e-mail: nanantha@umpa.ens-lyon.fr

Steve Zelditch
Johns Hopkins University
Baltimore, MD 21218, USA
e-mail: szelditch@jhu.edu

Communicated by Jens Marklof
Submitted: April 20, 2006
Revised: July 10, 2006
Accepted: July 31, 2006