

# ANNALES DE L'INSTITUT FOURIER

NIGEL J. E. PITT

## **On pairs of closed geodesics on hyperbolic surfaces**

*Annales de l'institut Fourier*, tome 49, n° 1 (1999), p. 1-25

[http://www.numdam.org/item?id=AIF\\_1999\\_\\_49\\_1\\_1\\_0](http://www.numdam.org/item?id=AIF_1999__49_1_1_0)

© Annales de l'institut Fourier, 1999, tous droits réservés.

L'accès aux archives de la revue « Annales de l'institut Fourier » (<http://annalif.ujf-grenoble.fr/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## ON PAIRS OF CLOSED GEODESICS ON HYPERBOLIC SURFACES

by Nigel J.E. PITT (\*)

---

### 1. Introduction.

Let  $\Gamma \backslash \mathfrak{H}$  be a compact, unramified hyperbolic surface; that is, a quotient of the upper half plane  $\mathfrak{H}$  by a Fuchsian group  $\Gamma < \mathrm{PSL}_2(\mathbb{R})$  with no parabolic or elliptic elements. Such a group is called totally hyperbolic. It is well known that the geometry of  $\Gamma \backslash \mathfrak{H}$  is intimately linked with the spectral properties of the hyperbolic Laplacian

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

as is apparent in the Selberg trace formula, which links the spectrum of  $\Delta$  with the lengths of closed geodesics on  $\Gamma \backslash \mathfrak{H}$  in a Poisson-like summation formula. Here we consider interactions between closed geodesics on  $\Gamma \backslash \mathfrak{H}$ , that is, the angles of intersection and lengths of common perpendiculars. The main result is another summation formula linking this information with spectral data for  $\Delta$ , and we use it to study these interactions both on average and individually.

We begin by outlining the trace formula, since this includes both notation and results we will need in the following. Let  $\rho(z, w)$  denote the hyperbolic distance between  $z, w \in \mathfrak{H}$ , and let

$$u(z, w) = e^{\rho(z, w)} + e^{-\rho(z, w)} - 2.$$

---

(\*) Research supported by CNPq-Brasil.

*Keywords:* Hyperbolic surfaces – Closed geodesics – Eigenfunctions of the Laplacian – Trace formula.

*Math. classification:* 11F72 – 53C22.

Let  $k$  be a compactly supported  $C^\infty$  test function on  $[0, \infty)$  and let  $K(z, w)$  denote the automorphic kernel

$$(1.1) \quad K(z, w) = \sum_{\gamma \in \Gamma} k(u(z, w)).$$

(In fact weaker conditions on  $k$  suffice in the following discussion, see [3], [10], but are not needed here). Let  $\{\varphi_j\}$  denote a complete set of eigenfunctions of the Laplacian in  $\mathcal{L}^2(\Gamma \backslash \mathfrak{H})$ , orthonormal with respect to the Petersson inner product

$$\langle f, g \rangle = \int_{\Gamma \backslash \mathfrak{H}} f(z) \bar{g}(z) d\mu z$$

for  $d\mu z$  the invariant measure  $y^{-2} dx dy$ , with

$$\Delta \varphi_j = \lambda_j \varphi_j \quad \text{for } \lambda_j = \frac{1}{4} + r_j^2 > 0.$$

The spectral theorem gives a decomposition of  $K$  as

$$(1.2) \quad K(z, w) = \sum_j h(r_j) \varphi_j(z) \bar{\varphi}_j(w)$$

where  $k(u)$  and  $h(r)$  are linked by the Selberg/Harish-Chandra transform, defined by

$$(1.3a) \quad h(r) = \int_{-\infty}^{\infty} G(\xi) e^{ir\xi} d\xi, \quad G(\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(r) e^{-ir\xi} dr,$$

$$(1.3b) \quad Q(e^\xi + e^{-\xi} - 2) = G(\xi),$$

$$(1.3c) \quad Q(w) = \int_w^\infty \frac{k(u)}{\sqrt{u-w}} du, \quad k(u) = -\frac{1}{\pi} \int_u^\infty \frac{dQ(w)}{\sqrt{w-u}},$$

where  $h$  is holomorphic in  $|\operatorname{Im} r| < \frac{1}{2} + \epsilon$ , is even in  $r$ , and decays as  $h(r) \ll (|r| + 1)^{-\delta}$  for  $\delta > 2$  (see [3], [10]). Note that there are finitely many complex  $r_j$ , corresponding to eigenvalues below  $\frac{1}{4}$ ; these include  $\lambda_0 = 0$  and possibly  $\lambda_1$ . Selberg's trace formula is now proved by evaluating the trace

$$(1.4) \quad \operatorname{Tr}_1(\Gamma, k) = \int_{\Gamma \backslash \mathfrak{H}} K(z, z) d\mu z$$

in two ways; the first using (1.2). The other uses a decomposition of  $\Gamma$  into conjugacy classes and an unfolding of the integral into regions where it can

be calculated. The conjugacy classes are naturally in 1–1 correspondence with the closed geodesics on  $\Gamma \backslash \mathfrak{H}$ , and the Jordan canonical form of the conjugacy class is a dilation  $z \mapsto pz$  where  $\log p$  is the length of the geodesic. We write

$$d_m(p) = |p^{m/2} - p^{-m/2}| \quad \text{and} \quad d(p) = d_1(p).$$

A geodesic is called *primitive* when it closes at the first opportunity, that is, the corresponding conjugacy class is not a power of another.

Evaluating using these ideas the Selberg trace formula in our case of totally hyperbolic groups can be written

$$(1.5) \quad \sum_P \log p \sum_{m \neq 0} \frac{Q(d_m^2(p))}{d_m(p)} = \sum_j h(r_j) - k(0) \text{Vol}(\Gamma \backslash \mathfrak{H})$$

where  $P$  ranges over primitive closed geodesics. This has been used both to give a Weyl law for the spectrum of  $\Delta$ ,

$$(1.6) \quad \#\{j : |r_j| \leq R\} \sim \frac{\text{Vol}(\Gamma \backslash \mathfrak{H})}{4\pi} R^2,$$

and to count primitive closed geodesics of restricted length:

$$(1.7) \quad \sum_{p \leq X} \log p = \sum_{0 < ir_j \leq \frac{1}{2}} \frac{X^{\frac{1}{2} + ir_j}}{\frac{1}{2} + ir_j} + O(X^{\frac{3}{4}}).$$

(Due to Selberg. Such a result is known as a prime geodesic theorem; see [2], [3], [5] and [8] for details, stronger versions, and connections with quadratic forms).

To introduce pairs of closed geodesics we use two variable versions of (1.2) and (1.3), providing a double sum over the group, and calculate a trace analogous to (1.4) to see the interactions between these pairs. The principal result of the paper is the following theorem:

**THEOREM 1.** — *Let  $\Gamma$  be a totally hyperbolic co-compact Fuchsian group, let  $\{\varphi_j\}$  denote a complete set of eigenfunctions of  $\Delta$  in  $\mathfrak{L}^2(\Gamma \backslash \mathfrak{H})$  with eigenvalues  $\lambda_j = \frac{1}{4} + r_j^2$ , and let  $\mathcal{P}$  denote the primitive closed geodesics on  $\Gamma \backslash \mathfrak{H}$ . Then the spectral expression*

$$\sum_{j,k} h(r_j, r_k) \langle |\varphi_j|^2, |\varphi_k|^2 \rangle - k(0) \text{Vol}(\Gamma \backslash \mathfrak{H})$$

is equal to the geometric expression

$$\begin{aligned} & \sum_{P \in \mathcal{P}} \log p \sum_{\substack{(m,n) \\ \neq (0,0)}} \frac{Q(d_m^2(p) + d_n^2(p))}{(d_m^2(p) + d_n^2(p))^{\frac{1}{2}}} \\ & + \sum_{P_1, P_2 \in \mathcal{P}} \sum_{\vartheta} \sum_{m, n \neq 0} \int_0^\infty \frac{Q(u + d_m^2(p_1) + d_n^2(p_2)) du}{\sqrt{u^3 + u^2(d_m^2(p_1) + d_n^2(p_2)) + u d_m^2(p_1) d_n^2(p_2) \sin^2 \vartheta}} \\ & + 2 \sum_{P_1, P_2 \in \mathcal{P}} \sum_{\rho} \sum_{m, n \neq 0} \int_A^\infty \frac{Q(u + d_m^2(p_1) + d_n^2(p_2)) du}{\sqrt{u^3 + u^2(d_m^2(p_1) + d_n^2(p_2)) - u d_m^2(p_1) d_n^2(p_2) \sinh^2 \rho}} \end{aligned}$$

where  $\vartheta$  ranges over the non-zero intersection angles between  $P_1$  and  $P_2$ ,  $\rho$  ranges over the geodesic common perpendiculars between  $P_1$  and  $P_2$ , and  $A$  denotes the largest root of the cubic expression. The test function  $Q$  is assumed to be  $C^\infty$  and of compact support<sup>(2)</sup>, and the various transforms are given by

$$k(u) = -\frac{1}{\pi} \int_u^\infty \frac{Q'(w) dw}{\sqrt{w-u}}, \quad L(w) = \int_w^\infty \frac{Q(u) du}{\sqrt{u-w}}, \quad Q(u) = -\frac{1}{\pi} \int_u^\infty \frac{L'(w) dw}{\sqrt{w-u}},$$

$$L(4 \sinh^2 \frac{1}{2} \xi_1 + 4 \sinh^2 \frac{1}{2} \xi_2) = G(\xi_1, \xi_2),$$

$$h(r_1, r_2) = \int_{-\infty}^\infty \int_{-\infty}^\infty G(\xi_1, \xi_2) e^{ir_1 \xi_1 + ir_2 \xi_2} d\xi_1 d\xi_2,$$

$$G(\xi_1, \xi_2) = \frac{1}{4\pi^2} \int_{-\infty}^\infty \int_{-\infty}^\infty h(r_1, r_2) e^{-ir_1 \xi_1 - ir_2 \xi_2} dr_1 dr_2.$$

On the geometric side the new information here is encoded in the structure of the integrals appearing. These are of elliptic type, becoming unbounded as  $\vartheta \rightarrow 0$  or  $\rho \rightarrow 0$  in typical choices of  $Q$ . On average, then, the theorem restricts the degree to which very small intersection angles or very short common perpendiculars can occur, since the spectral side can be evaluated asymptotically for “nice” choices of  $Q$ . The complication is that the integrals depend not only on  $\vartheta$  or  $\rho$ , but on a function of the lengths of the closed geodesics and the interaction. This is to be expected, as can be seen by considering the intersection angles of closed geodesics in the analogous case of the torus  $\mathbb{Z}^2 \backslash \mathbb{R}^2$ . Closed geodesics on  $\mathbb{Z}^2 \backslash \mathbb{R}^2$  occur

<sup>(2)</sup> The restriction to compact support can surely be weakened to some polynomial decay of  $Q$  and its first derivatives, but this is not necessary for our applications here.

in parallel families, coming from straight lines in the plane with rational slope. If this slope in reduced form is  $p/q$ , say, then the primitive closed geodesic associated to it has length  $(p^2 + q^2)^{\frac{1}{2}}$ , and as can be seen from the geometry of the situation, two such geodesics corresponding to slopes of  $p_1/q_1$  and  $p_2/q_2$  will intersect  $|p_1q_2 - q_1p_2|$  times, always at the same angle  $\vartheta$ , given by

$$\sin \vartheta = \frac{|p_1q_2 - q_1p_2|}{(p_1^2 + q_1^2)^{\frac{1}{2}}(p_2^2 + q_2^2)^{\frac{1}{2}}}.$$

Thus a small angle of intersection requires that at least one of the geodesics is long. The negative curvature, non-abelian case of  $\Gamma \backslash \mathfrak{H}$  is more complex; the lengths of individual primitive closed geodesics are not simply calculated, and two primitive closed geodesics will probably not repeat an intersection angle, nonetheless Theorem 1 controls the interactions on average.

In the case of intersection points the moduli of the elliptic integrals has a geometric interpretation. Geodesic polar coordinates  $(\rho, \varphi)$  are defined for  $\mathfrak{H}$  (see [3]) by considering any  $z \in \mathfrak{H}$  as lying on a unique geodesic passing through  $i$  at an angle  $\varphi \in [0, 2\pi)$  from the vertical, at a distance  $\rho \in [0, \infty)$  from  $i$ . Linking  $\rho$  with  $u$  as above we can construct a tangent plane to  $\mathfrak{H}$  at  $i$  as the euclidean plane with polar coordinates  $(\sqrt{u}, \varphi)$ . By translation this clearly also defines a tangent plane at an arbitrary  $w \in \mathfrak{H}$ , which descends naturally to a tangent plane to  $\Gamma \backslash \mathfrak{H}$  under the covering map. If two closed geodesics  $P_1$  and  $P_2$  cross at  $z$  at an angle  $\vartheta$ , then they are the images under the covering map  $\mathfrak{H} \rightarrow \Gamma \backslash \mathfrak{H}$  of two geodesic segments in  $\mathfrak{H}$  of lengths  $\log p_1$  and  $\log p_2$  respectively, crossing at their midpoints at the same angle  $\vartheta$ . The corresponding trajectories in the tangent space are euclidean line segments of lengths  $d(p_1), d(p_2)$  respectively, crossing at their midpoints at an angle of  $\vartheta$ . There is a unique euclidean ellipse (see Fig. 1) passing all four endpoints parallel to the other line segment, and the modulus of the integral appearing in Theorem 1 is precisely the eccentricity of this ellipse.

**DEFINITION.** — *For each intersection point of closed geodesics of lengths  $\log p$  and  $\log q$  we define the eccentricity  $\kappa \in [0, 1)$  of the intersection as being the eccentricity of the corresponding ellipse in the tangent space.*

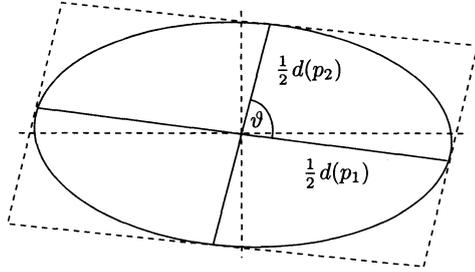


Figure 1

This is given by

$$\kappa^2 = \frac{2\sqrt{d^4(p_1) + 2d^2(p_1)d^2(p_2) \cos 2\vartheta + d^4(p_2)}}{d^2(p_1) + d^2(p_2) + \sqrt{d^4(p_1) + 2d^2(p_1)d^2(p_2) \cos 2\vartheta + d^4(p_2)}}$$

where  $\vartheta$  is the angle of intersection. Let  $\mathcal{K}_1(P_1, P_2)$  denote the set of such eccentricities. For each common perpendicular of length  $\rho$  we define the eccentricity  $\kappa \in [1/\sqrt{2}, 1)$  by analogy:

$$\kappa^2 = \frac{d^2(p_1) + d^2(p_2) + \sqrt{d^4(p_1) + 2d^2(p_1)d^2(p_2) \cosh 2\rho + d^4(p_2)}}{2\sqrt{d^4(p_1) + 2d^2(p_1)d^2(p_2) \cosh 2\rho + d^4(p_2)}}$$

and let  $\mathcal{K}_2(P_1, P_2)$  denote the set of such eccentricities.

With these definitions we have our first corollary, which is essentially an extraction of an asymptotic form of Theorem 1.

COROLLARY 1. — Let  $F$  denote the elliptic integral of the first kind

$$F(\varphi, \kappa) = \int_0^\varphi \frac{d\alpha}{\sqrt{1 - \kappa^2 \sin^2 \alpha}}$$

and let  $\rho_1$  be  $|r_1|$  if  $\lambda_1 = \frac{1}{4} + r_1^2 < \frac{1}{4}$ , 0 otherwise.

Then for  $X \gg 1$  we have

$$\begin{aligned}
 & \sum_{d^2(p_1)+d^2(p_2) \leq X} \sum_{\kappa \in \mathcal{K}_1(P_1, P_2)} \frac{2^{\frac{1}{2}}(2-\kappa^2)^{\frac{1}{2}}}{\kappa(d^2(p_1)+d^2(p_2))^{\frac{1}{2}}} \\
 & \quad \times F\left(\tan^{-1} \sqrt{\frac{2-\kappa^2}{1-\kappa^2} \left(\frac{X}{d^2(p_1)+d^2(p_2)} - 1\right)}, \kappa\right) \\
 & + \sum_{d^2(p_1)+d^2(p_2) \leq X} \sum_{\kappa \in \mathcal{K}_2(P_1, P_2)} \frac{4(2\kappa^2-1)^{\frac{1}{2}}}{(d^2(p_1)+d^2(p_2))^{\frac{1}{2}}} \\
 & \quad \times F\left(\sec^{-1} \sqrt{\frac{2\kappa^2-1}{1-\kappa^2} \left(\frac{X}{d^2(p_1)+d^2(p_2)} - 1\right)}, \kappa\right) \\
 & = \frac{2\pi X^{\frac{3}{2}}}{\text{Vol}(\Gamma \backslash \mathfrak{H})} + O(X^{\rho_1+1} + X^{\frac{29}{22}+\epsilon})
 \end{aligned}$$

where the summations are over all closed geodesics satisfying the restriction, and the integral in the second sum is understood to vanish if  $\kappa$  is such that  $\sec^{-1}$  is not defined.

Note that as remarked above, if  $\vartheta$  or  $\rho$  tends to 0 then the integrals tend to the complete elliptic integral of modulus 1, which is divergent. Corollary 1 can thus be seen as restricting the average behaviour of the interactions. It is interesting to speculate, however, that the arguments of the integrals should have some geometric and/or physical interpretation related to the global structure of the sum, although we make no progress in this direction here.

To consider an individual pair of geodesics it is more efficient to return to Theorem 1 using a different choice of test function  $Q$ , noting that an interaction of a general pair of closed geodesics must also be an interaction of a pair of primitive such. We prove the following corollary:

**COROLLARY 2.** — *Let  $P_1$  and  $P_2$  be two primitive closed geodesics. If  $\alpha$  denotes any constant greater than  $\frac{16}{9}$ , and  $c$  is any positive constant greater than  $2\sqrt{2}\pi/\text{Vol}(\Gamma \backslash \mathfrak{H})$ , then for sufficiently large  $d^2(p_1) + d^2(p_2)$  we have*

$$\begin{aligned}
 \sum_{\vartheta} \sinh^{-1} \left( \frac{(d^2(p_1) + d^2(p_2))^\alpha}{d^2(p_1)d^2(p_2)\sin^2 \vartheta} \right)^{\frac{1}{2}} + \sum_{\rho} \cosh^{-1} \left( \frac{(d^2(p_1) + d^2(p_2))^\alpha}{d^2(p_1)d^2(p_2)\sinh^2 \rho} \right)^{\frac{1}{2}} \\
 \leq c(d^2(p_1) + d^2(p_2))^\alpha
 \end{aligned}$$

where  $\vartheta$  and  $\rho$  vary over all angles of intersection and common perpendiculars, which individually can be bounded below by

$$\sin \vartheta \geq \frac{(d^2(p_1) + d^2(p_2))^{\frac{\alpha}{2}}}{d(p_1)d(p_2) \sinh(c(d^2(p_1) + d^2(p_2))^\alpha)},$$

$$\sinh \rho \geq \frac{(d^2(p_1) + d^2(p_2))^{\frac{\alpha}{2}}}{d(p_1)d(p_2) \cosh(c(d^2(p_1) + d^2(p_2))^\alpha)}.$$

Note that this result says nothing about short closed geodesics, and indeed is in some sense weaker than Corollary 1, since statements about individual cases have been deduced from an average. In fact if one wishes to consider only individual  $\vartheta$  or  $\rho$  then the lower bounds here are not as strong as can be proved by other methods (see [1]).

The approach to these results, loosely speaking, is the following. Given a function  $k_1(u_1, u_2)$  of compact support in  $[0, \infty) \times [0, \infty)$ , a two-variable version of the Selberg/Harish-Chandra transform can be defined by

$$Q_1(w_1, w_2) = \int_{w_2}^{\infty} \int_{w_1}^{\infty} \frac{k_1(u_1, u_2) du_1 du_2}{\sqrt{(u_1 - w_1)(u_2 - w_2)}}$$

$$G_1(\xi_1, \xi_2) = Q_1(e^{\xi_1} + e^{-\xi_1} - 2, e^{\xi_2} + e^{-\xi_2} - 2)$$

$$h_1(r_1, r_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_1(\xi_1, \xi_2) e^{ir_1\xi_1 + ir_2\xi_2} d\xi_1 d\xi_2$$

and the expression

$$\mathrm{Tr}_2(\Gamma, k_1) = \int_{\Gamma \backslash \mathfrak{H}} \sum_{g_1, g_2 \in \Gamma} k_1(u(z, g_1 z), u(z, g_2 z)) d\mu z$$

can be expanded both spectrally and geometrically. This initially produces a more general but less convenient theorem, a special case of which was discussed in [6].

**THEOREM 2.** — *Let  $\Gamma, \{\varphi_j\}, \mathcal{P}, \vartheta, \rho$  be as above, and let  $h_1(r_1, r_2), k_1(u_1, u_2)$  be a two-variable Selberg/Harish-Chandra pair,  $k_1$  of compact support. Then the spectral expression*

$$\sum_{j,k} h_1(r_j, r_k) \langle |\varphi_j|^2, |\varphi_k|^2 \rangle - k_1(0, 0) \mathrm{Vol}(\Gamma \backslash \mathfrak{H})$$

is equal to the geometric expression

$$\begin{aligned} & \sum_{P \in \mathcal{P}} \log p \sum_{\substack{(m,n) \\ \neq (0,0)}} \int_{-\infty}^{\infty} k_1(d_m^2(p)(x^2+1), d_n^2(p)(x^2+1)) dx \\ & + \sum_{P_1, P_2 \in \mathcal{P}} \sum_{\vartheta} \sum_{m, n \neq 0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{k_1(d_m^2(p_1)(t_1^2+1), d_n^2(p_2)(t_2^2+1)) dt_1 dt_2}{\sqrt{t_1^2 + t_2^2 + 2 \cos \vartheta t_1 t_2 + \sin^2 \vartheta}} \\ & + 2 \sum_{P_1, P_2 \in \mathcal{P}} \sum_{\rho} \sum_{m, n \neq 0} \int_{-\infty}^{\infty} \int_{B(t_2)} \frac{k_1(d_m^2(p_1)(t_1^2+1), d_n^2(p_2)(t_2^2+1)) dt_1 dt_2}{\sqrt{t_1^2 + t_2^2 + 2 \cosh \rho t_1 t_2 - \sinh^2 \rho}} \end{aligned}$$

where  $B(t_2)$  denotes the larger root of the radical.

While Theorem 2 is more general than Theorem 1 in terms of the test functions allowed, it is harder to produce quantitative results through choices of  $k_1$  due to the double integrals appearing. Theorem 1 is proved from Theorem 2 by using  $k_1(u_1, u_2) = k(u_1 + u_2)$ , so that  $Q_1(w_1, w_2) = L(w_1 + w_2)$ , where

$$(1.8) \quad L(w) = \int_w^{\infty} \frac{Q(u) du}{\sqrt{u-w}}, \quad Q(u) = -\frac{1}{\pi} \int_u^{\infty} \frac{L'(w) dw}{\sqrt{w-u}}$$

and  $h(r_1, r_2)$ ,  $L(w)$  are related by the Fourier pair as in Theorem 1.

It is worth noting that an apparent approach to these theorems would be to expand one of the automorphic kernels into spectral data and to use the more usual conjugation of  $\Gamma$  to expand the second. In fact this method does not seem to work well since the spectral sum remaining must be treated geometrically also, and it is not as simple to do this a second time. While this can surely be done in principle the route appears more complicated. Note however the work of Zelditch [11], where the trace is of an automorphic kernel against a single eigenfunction, and produces integrals of this eigenfunction along closed geodesics.

## 2. The spectral trace.

Using (1.2) for both sums over the group we have an expansion

$$\begin{aligned} & \sum_{g_1, g_2 \in \Gamma} k_1(u(z_1, g_1 w_1), u(z_2, g_2 w_2)) \\ & = \sum_{j, k} h_1(r_j, r_k) \varphi_j(z_1) \bar{\varphi}_j(w_1) \varphi_k(z_2) \bar{\varphi}_k(w_2) \end{aligned}$$

which is absolutely and uniformly convergent. Term-by-term integration gives

$$(2.1) \quad \begin{aligned} \mathrm{Tr}_2(\Gamma, k_1) &= \int_{\Gamma \backslash \mathfrak{H}} \sum_{g_1, g_2 \in \Gamma} k_1(u(z, g_1 z), u(z, g_2 z)) \, d\mu z \\ &= \sum_{j, k} h_1(r_j, r_k) \langle |\varphi_j|^2, |\varphi_k|^2 \rangle \end{aligned}$$

giving the main expression on the spectral side of Theorem 2.

### 3. Decomposition into conjugacy classes.

Each of the two sums over group elements includes the identity; considering this separately in each sum by inclusion and exclusion and noting that  $u(z, z) = 0$  we obtain

$$(3.1) \quad \begin{aligned} \mathrm{Tr}_2(\Gamma, k_1) &= \int_{\Gamma \backslash \mathfrak{H}} \sum_{g_1, g_2 \neq I} k_1(u(z, g_1 z), u(z, g_2 z)) \, d\mu z - k_1(0, 0) \mathrm{Vol}(\Gamma \backslash \mathfrak{H}) \\ &\quad + \int_{\Gamma \backslash \mathfrak{H}} \sum_g k_1(0, u(z, g z)) \, d\mu z + \int_{\Gamma \backslash \mathfrak{H}} \sum_g k_1(u(z, g z), 0) \, d\mu z. \end{aligned}$$

The first integral, which we denote  $\mathrm{Tr}_2^*(\Gamma, k_1)$  has all the new information since the last two terms can (and will) be considered using the Selberg trace formula (1.5).

Consider conjugation of  $\Gamma \times \Gamma - \{(I, I)\}$  by  $\Gamma$ :

$$\tau^{-1}(g_1, g_2)\tau = (\tau^{-1}g_1\tau, \tau^{-1}g_2\tau), \quad \tau, g_1, g_2 \in \Gamma.$$

This decomposes  $\Gamma \times \Gamma - \{(I, I)\}$  into conjugacy classes  $\{(g_1, g_2)\}$ , for which  $\tau_1, \tau_2 \in \Gamma$  give the same conjugate of  $(g_1, g_2)$  if and only if

$$(\tau_1^{-1}g_1\tau_1, \tau_1^{-1}g_2\tau_1) = (\tau_2^{-1}g_1\tau_2, \tau_2^{-1}g_2\tau_2).$$

This is equivalent with  $\tau_1\tau_2^{-1}$  commuting with both  $g_1$  and  $g_2$ , in other words  $\tau_1\tau_2^{-1}$  is an element of the joint centraliser  $\Gamma(g_1, g_2)$ . Thus from the point-pair invariance of  $u$  we obtain

$$\begin{aligned} \mathrm{Tr}_2^*(\Gamma, k_1) &= \sum_{\{(g_1, g_2)\}}^* \sum_{\tau \in \Gamma(g_1, g_2) \backslash \Gamma} \int_{\Gamma \backslash \mathfrak{H}} k_1(u(z, \tau^{-1}g_1\tau z), u(z, \tau^{-1}g_2\tau z)) \, d\mu z \\ &= \sum_{\{(g_1, g_2)\}}^* \int_{\Gamma(g_1, g_2) \backslash \mathfrak{H}} k_1(u(z, g_1 z), u(z, g_2 z)) \, d\mu z \end{aligned}$$

where  $*$  indicates that neither  $g_1$  nor  $g_2$  is the identity.

There are now distinct situations to be considered. If  $g_1$  and  $g_2$  have the same fixed points then  $\Gamma(g_1, g_2)$  is simply the centraliser  $\Gamma(g_1)$  corresponding to a primitive element with the same fixed set. If they do not, then no non-trivial element commutes with both and the integration is over all of  $\mathfrak{H}$ . We analyse the integral accordingly in the following sections.

#### 4. The case of equal fixed points.

This case is very similar to that in the proof of the trace formula, since if  $g_1$  and  $g_2$  have the same fixed points then they are both powers of the same primitive element. Representatives of all such conjugacy classes can now be chosen as  $(P^m, P^n)$  where  $P$  varies over primitive conjugacy classes as above and  $m, n \neq 0$ . If  $P$  has fixed points  $\alpha$  and  $\beta$  then we can choose  $T \in \text{PSL}_2(\mathbb{R})$  to map these to 0 and  $\infty$ , by  $T: z \mapsto (z - \alpha)/(z - \beta)$  or a translation in the case one is already  $\infty$ . In either case  $T^{-1}P^mT$  maps  $z \mapsto p^{\pm m}z$ , and using a variable change  $z \mapsto Tz$  in the integral and the point-pair invariance of  $u$  the total contribution from all such classes is

$$(4.1) \quad \sum_P \sum_{m, n \neq 0} \int_1^p \int_{-\infty}^{\infty} k_1(u(z, p^m z), u(z, p^n z)) d\mu.$$

Using

$$(4.2) \quad u(z, w) = \frac{|z - w|^2}{\text{Im } z \text{ Im } w}$$

and  $d_m(p)$  as above, a direct calculation now gives

$$(4.3) \quad \sum_P \log p \sum_{m, n \neq 0} \int_{-\infty}^{\infty} k_1(d_m^2(p)(x^2 + 1), d_n^2(p)(x^2 + 1)) dx.$$

#### 5. The case of differing fixed points.

Two hyperbolic elements of a Fuchsian group sharing a fixed point must share both (see [4]), so if  $g_1$  and  $g_2$  do not have the same fixed pair they must have fixed points  $\alpha_1, \beta_1$  and  $\alpha_2, \beta_2$  respectively, all four mutually distinct. Thus the contribution from all cases not handled in §3 is

$$(5.1) \quad \sum_{\{(g_1, g_2)\}}^* \int_{\mathfrak{H}} k_1(u(z, g_1 z), u(z, g_2 z)) d\mu z$$

where  $*$  indicates that  $g_1$  and  $g_2$  share no fixed point.

LEMMA 5.1. — *Representatives for the conjugacy classes  $\{(g_1, g_2)\}$  can be chosen to be  $(P_1, \tau^{-1}P_2\tau)$  where  $P_1, P_2 \in \mathcal{P}$  and  $\tau$  ranges over elements of the double coset  $\Gamma(P_2)\backslash\Gamma/\Gamma(P_1)$ , where  $\Gamma(g)$  denotes the centraliser of  $g$  in  $\Gamma$ .*

*Proof.* — First conjugate the pairs to give  $(P_1, g_2)$  in each case, for some  $P_1 \in \mathcal{P}$ , where  $g_2 \in \Gamma$  can be written uniquely as  $\tau^{-1}P_2\tau$  for some  $P_2 \in \mathcal{P}$  and  $\tau \in \Gamma(P_2)\backslash\Gamma$ . Thus two pairs  $(P_1, \tau_1^{-1}P_2\tau_1), (P_3, \tau_2^{-1}P_4\tau_2)$  are diagonally conjugate if and only if  $P_1 = P_3, P_2 = P_4$ , and  $\tau_2 = \tau_1\sigma$  for some  $\sigma \in \Gamma(P_1)$ .

Using this lemma (5.1) can be written as

$$(5.2) \quad \sum_{P_1, P_2} \sum_{m, n \neq 0} \sum_{\tau \in \Gamma(P_2)\backslash\Gamma/\Gamma(P_1)}^* \int_{\mathfrak{H}} k_1(u(z, P_1^m z), u(z, \tau^{-1}P_2^n \tau z)) \, d\mu z$$

where  $P_1$  and  $P_2$  are as above and  $*$  indicates that  $\tau^{-1}P_2\tau$  does not have the same fixed points as  $P_1$ , which is equivalent to saying that  $\tau$  is not the identity if  $P_1 = P_2$ . Note that this expression is symmetric in  $P_1$  and  $P_2$  by a variable change in the integral and the point-pair invariance of  $u$ . The sums over  $P_1^m, P_2^n$  can be viewed as being over closed geodesics on  $\Gamma\backslash\mathfrak{H}$ , with  $m = n = 1$  being the case where both are primitive; however we still require a geometric interpretation of the remaining sum over  $\tau$ .

## 6. Cross-ratio and pairs of geodesics.

Here the terms “image” and “preimage” refer to the covering map  $\mathfrak{H} \rightarrow \Gamma\backslash\mathfrak{H}$ . Let  $\gamma_g$  denote the geodesic in  $\mathfrak{H}$  corresponding to  $g \in \Gamma$ , and let  $\gamma_g^*$  denote the associated closed geodesic on  $\Gamma\backslash\mathfrak{H}$ . For each pair  $P_1, P_2$  we consider the sets

$$\begin{aligned} \mathcal{S}_1 &= \{\tau \in \Gamma(P_1)\backslash\Gamma/\Gamma(P_2) : \tau\gamma_{P_2} \cap \gamma_{P_1} \neq \emptyset\}, \\ \mathcal{S}_2 &= \{\tau \in \Gamma(P_1)\backslash\Gamma/\Gamma(P_2) : \tau\gamma_{P_2} \cap \gamma_{P_1} = \emptyset\} \end{aligned}$$

and will show that these correspond respectively to intersections and common perpendiculars of  $\gamma_{P_1}^*$  and  $\gamma_{P_2}^*$ . Note that any element of  $\mathcal{S}_1$  defines a unique point  $z$  in  $\mathfrak{H}$ , and that any element of  $\mathcal{S}_2$  defines a unique geodesic segment in  $\mathfrak{H}$  meeting  $\gamma_{P_1}$  and  $\tau\gamma_{P_2}$  orthogonally.

Consider  $\mathcal{S}_1$ . Any intersection  $z^*$  of  $\gamma_{P_1}^*$  and  $\gamma_{P_2}^*$  is the image of an intersection of some translate of  $\gamma_{P_1}$  and some translate of  $\gamma_{P_2}$  by elements

of  $\Gamma$ , as can be seen by lifting a neighbourhood of  $z^*$  to  $\mathfrak{H}$  and extending the geodesic segments to complete geodesics in  $\mathfrak{H}$ : the resulting geodesics descend to  $\gamma_{P_1}^*$  and  $\gamma_{P_2}^*$  and are thus translates of  $\gamma_{P_1}$  and  $\gamma_{P_2}$  as claimed. By a translation of  $\mathfrak{H}$  we may suppose the translate of  $\gamma_{P_1}$  is  $\gamma_{P_1}$  itself, or in other words every intersection  $z^*$  of  $\gamma_{P_1}^*$  and  $\gamma_{P_2}^*$  is the image of an intersection of  $\gamma_{P_1}$  and some translate of  $\gamma_{P_2}$ . Observe now that two geodesics  $\tau_1\gamma_g, \tau_2\gamma_g$  for  $\tau_1, \tau_2 \in \Gamma$  are the same set-wise if and only if  $\tau_1$  and  $\tau_2$  are equivalent under the action of  $\Gamma(g)$  on the right, so

$$G_g = \{\tau\gamma_g : \tau \in \Gamma/\Gamma(g)\} = \{\gamma_{\tau g \tau^{-1}} : \tau \in \Gamma/\Gamma(g)\}$$

is the set of all the set-wise distinct translates of  $\gamma_g$ . Two intersections of  $\gamma_{P_1}$  with elements of  $G_{P_2}$  clearly descend to the same intersection point on  $\Gamma \backslash \mathfrak{H}$  if they differ only by an element of  $\Gamma(P_1)$  on the left, so we now have a well-defined surjective map

$$\mathcal{S}_1 \longrightarrow \{z_1^*, z_2^*, \dots\}$$

where  $z_i^*$  are the intersections of  $\gamma_{P_1}^*$  and  $\gamma_{P_2}^*$ . This map can also be seen to be injective, since intersections of  $\tau_1\gamma_{P_2}$  and  $\tau_2\gamma_{P_2}$  with  $\gamma_{P_1}$  are equivalent under  $\Gamma$  if and only if they are equivalent under  $\Gamma(P_1)$ . Thus we have the following lemma:

LEMMA 6.1. — *Let  $P_1$  and  $P_2$  be primitive hyperbolic elements of  $\Gamma$ , perhaps equal. The points of intersection of  $\gamma_{P_1}^*$  and  $\gamma_{P_2}^*$ , counted with multiplicity, are in one-one correspondence with the geodesics  $\tau\gamma_{P_2}$  for  $\tau \in \Gamma(P_1) \backslash \Gamma/\Gamma(P_2)$  which cut  $\gamma_{P_1}$ .*

Consider now  $\mathcal{S}_2$ . If  $\ell^*$  of length  $\rho$  is a common perpendicular of  $\gamma_{P_1}^*$  and  $\gamma_{P_2}^*$  then we may lift  $\ell^*$  and a neighbourhood of it in  $\Gamma \backslash \mathfrak{H}$  (including the base points) to a geodesic segment  $\ell$  of length  $\rho$  and a neighbourhood of it in  $\mathfrak{H}$ . The preimages of the segments of  $\gamma_{P_1}^*$  and  $\gamma_{P_2}^*$  may be continued to give translates of  $\gamma_{P_1}$  and  $\gamma_{P_2}$ , and as above we may translate  $\mathfrak{H}$  to suppose that the translate of  $\gamma_{P_1}$  is  $\gamma_{P_1}$  itself. By noting that if two translates of  $\gamma_{P_2}$  differ only by an element of  $\Gamma(P_1)$  on the left then the common perpendiculars descend to the same common perpendicular on  $\Gamma \backslash \mathfrak{H}$  we now have a surjective map

$$\mathcal{S}_2 \longrightarrow \{\ell_1^*, \ell_2^*, \dots\}.$$

The injectivity follows as above, thus the following lemma:

LEMMA 6.2. — *Let  $P_1$  and  $P_2$  be primitive hyperbolic elements of  $\Gamma$ , perhaps equal. The elements of  $\Gamma(P_1)\backslash\Gamma/\Gamma(P_2)$  such that  $\tau\gamma_{P_2}$  does not cut  $\gamma_{P_1}$  are in 1–1 correspondence with the geodesic paths (not necessarily closed) on  $\Gamma\backslash\mathfrak{H}$  which meet both  $\gamma_{P_1}^*$  and  $\gamma_{P_2}^*$  orthogonally.*

Let us return now to the integral in (5.2),

$$\int_{\mathfrak{H}} k_1(u(z, g_1 z), u(z, g_2 z)) d\mu z$$

where  $g_1$  and  $g_2$  have distinct endpoints. The well-known cross-ratio of four points on  $\mathbb{R} \cup \{\infty\}$  is defined by

$$[z_1, z_2; z_3, z_4] = \frac{z_3 - z_1}{z_2 - z_1} / \frac{z_4 - z_1}{z_2 - z_4}$$

and we will speak of the cross-ratio of  $\gamma_{g_1}$  and  $\gamma_{g_2}$  as being

$$\xi = [\alpha_1, \beta_1; \alpha_2, \beta_2]$$

where  $\alpha_i, \beta_i$  are the four endpoints as above. The four points are distinct so that the cross-ratio is neither zero nor infinite, but lacking a convention as to which of the endpoints is larger this gives a valid definition only up to the ambiguity between  $\xi$  and  $\xi^{-1}$ . However it will be positive or negative as the geodesics do not, or do cross, respectively. As in §3 we may shift  $z$ , and hence conjugate  $g_1$  and  $g_2$ , to assume without loss of generality that the fixed points of  $g_1$  are 0 and  $\infty$ . The fixed points of  $g_2$  are moved, but the cross-ratio of the four points has not changed. Using a second conjugation by a dilation  $z \mapsto tz$  if necessary we may now assume that the fixed points of  $g_2$  are 1 and  $\xi$ , and in the case where the geodesics do not cross we may further assume that  $\xi < 1$ . Thus we have one of the two situations shown (Fig. 2) below.

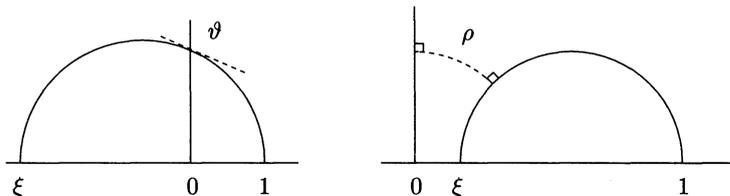


Figure 2

The geodesics have either a well-defined angle of intersection  $\vartheta$  or a unique common perpendicular of length  $\rho$ , which are related to  $\xi$  by

$$\vartheta = \cos^{-1} \left( \frac{1 + \xi}{1 - \xi} \right), \quad \rho = \cosh^{-1} \left( \frac{1 + \xi}{1 - \xi} \right)$$

respectively. The ambiguity between  $\xi$  and  $\xi^{-1}$  can now be seen to be insignificant; the two geodesics appear once from the pair  $(g_1, g_2)$  and a second time from the pair  $(g_2, g_1)$ , so in the first case both  $\vartheta$  and  $\pi - \vartheta$  appear, corresponding to  $\xi$  and  $\xi^{-1}$ , furthermore the integrals that appear are independent of this distinction. In the second case the length of the common perpendicular is the same for  $(g_1, g_2)$  and  $(g_2, g_1)$ .

The integral is now

$$\int_{\mathfrak{H}} k_1 \left( u(z, p_1^m z), u \left( z, \begin{pmatrix} 1 & \xi \\ 1 & 1 \end{pmatrix} \begin{pmatrix} p_2^{\frac{n}{2}} & \\ & p_2^{-\frac{n}{2}} \end{pmatrix} \begin{pmatrix} 1 & -\xi \\ -1 & 1 \end{pmatrix} z \right) \right) d\mu z$$

where  $p_1, p_2$  are the dilations associated to  $P_1$  and  $P_2$  in (4.2), as can be seen by considering the fixed points of the two matrices. Calculating the  $u$  functions using (3.2) now gives

$$\int_{\mathfrak{H}} k_1 \left( d_m^2(p_1) \frac{|z^2|}{y^2}, \frac{d_n^2(p_2)}{(1 - \xi)^2} \frac{|z - 1|^2 \cdot |z - \xi|^2}{y^2} \right) d\mu z.$$

This integral can be given another form, which makes the behaviour of  $d_m^2(p_1), d_n^2(p_2)$  and  $\xi$  more transparent in a qualitative sense. In the first case of crossing geodesics let

$$t_1 = \frac{x}{y}, \quad t_2 = \frac{(x - 1)(x - \xi) + y^2}{(1 - \xi)y},$$

the first being the slope of an euclidean line through 0, and the second being the quotient of the real and imaginary parts of  $(z - 1)/(z - \xi)$ , the analogous quantity for euclidean circles through 1 and  $\xi$ . Since  $x = yt_1$  we have

$$y^{-2} \frac{\partial(x, y)}{\partial(t_1, t_2)} = y^{-1} \frac{\partial y}{\partial t_2} = \frac{1 - \xi}{\sqrt{((1 + \xi)t_1 + (1 - \xi)t_2)^2 - 4\xi(1 + t_1^2)}}$$

and the integral becomes

$$\begin{aligned} (1 - \xi) \int_{\mathbb{R}^2} \frac{k_1(d_m^2(p_1)(t_1^2 + 1), d_n^2(p_2)(t_2^2 + 1))}{\sqrt{(1 - \xi)^2(t_1^2 + t_2^2) + 2(1 - \xi^2)t_1 t_2 - 4\xi}} dt_1 dt_2 \\ = \int_{\mathbb{R}^2} \frac{k_1(d_m^2(p_1)(t_1^2 + 1), d_n^2(p_2)(t_2^2 + 1))}{\sqrt{t_1^2 + t_2^2 + 2t_1 t_2 \cos \vartheta + \sin^2 \vartheta}} dt_1 dt_2. \end{aligned}$$

The case of non-intersecting geodesics can be calculated in a similar manner, but the variable change requires a division of the integral into two parts, giving the expression (and factor of 2) in the theorem.

Collecting together (2.1), (3.1), (4.3), (5.2) and applying the results of this section for the more complicated expressions we have a preliminary version of Theorem 2. The final form is now proved by noting that the third and fourth terms of (3.1) supply, by the Selberg trace formula (1.5), the degenerate cases  $m = 0, n \neq 0$  and  $m \neq 0, n = 0$  of the expression in (4.3), together with  $2k(0, 0)\text{Vol}(\Gamma \backslash \mathfrak{H})$ .

## 7. Proof of Theorem 1.

Theorem 1 will follow now from Theorem 2 by using the choice

$$k_1(u_1, u_2) = k(u_1 + u_2),$$

where  $k(u)$  is  $C^\infty$  and compactly supported on  $[0, \infty)$ , so the double Selberg/Harish transform  $h(r_1, r_2)$  exists, and has the form given in the theorem, as can be seen from (1.8).

The first integral appearing on the geometric side of Theorem 2 can be reduced to that in Theorem 1 by a direct application of (1.3). The second two are rather more complicated. Using

$$d_m(p_1)t_1 = \sqrt{u} \cos \varphi, \quad d_n(p_2)t_2 = \sqrt{u} \sin \varphi$$

we have

$$\begin{aligned} & \int_{\mathbb{R}^2} \frac{k(d_m^2(p_1)(t_1^2 + 1) + d_n^2(p_2)(t_2^2 + 1)) dt_1 dt_2}{\sqrt{t_1^2 + t_2^2 - 2 \cos \vartheta t_1 t_2 + \sin^2 \vartheta}} \\ &= \frac{1}{2d_m(p_1)d_n(p_2)} \int_0^{2\pi} (uF(\varphi) + \sin^2 \vartheta)^{-\frac{1}{2}} \int_0^\infty k(u + d_m^2(p_1) + d_n^2(p_2)) du d\varphi \end{aligned}$$

where

$$\begin{aligned} F(\varphi) = d_m^{-2}(p_1)d_n^{-2}(p_2) & \left( d_m^2(p_1) \sin^2 \varphi \right. \\ & \left. - 2d_m(p_1)d_n(p_2) \sin \varphi \cos \varphi \cos \vartheta + d_n^2(p_2) \cos^2 \varphi \right). \end{aligned}$$

Using (1.3c) and evaluating in  $u$  this is

$$\begin{aligned} & -\frac{1}{2\pi d_m(p_1)d_n(p_2)} \int_0^{2\pi} F^{-\frac{1}{2}}(\varphi) \int_0^\infty Q'(t + d_m^2(p_1) + d_n^2(p_2)) \\ & \quad \times \cos^{-1}\left(\frac{\sin^2 \vartheta - tF(\varphi)}{\sin^2 \vartheta + tF(\varphi)}\right) dt d\varphi \\ & = \frac{|\sin \vartheta|}{2\pi d_m(p_1)d_n(p_2)} \int_0^\infty Q(t + d_m^2(p_1) + d_n^2(p_2)) t^{-\frac{1}{2}} \\ & \quad \int_0^{2\pi} (\sin^2 \vartheta + tF(\varphi))^{-1} d\varphi dt \end{aligned}$$

using integration by parts in  $t$ ; the boundary terms vanish, due to the compact support of  $Q$ . Dividing the inner integral in four parts and using  $\tau = \tan \varphi$  we now obtain the integral as stated in Theorem 1.

The same variable change simplifies the last integral:

$$\begin{aligned} & \int_{-\infty}^\infty \int_{\substack{\sinh \rho(t_1^2+1)^{\frac{1}{2}} \\ -t_1 \cosh \rho}}^{\infty} \frac{k(d_m^2(p_1)(t_1^2+1) + d_n^2(p_2)(t_2^2+1))}{\sqrt{t_1^2 + t_2^2 + 2t_1t_2 \cosh \rho - \sinh^2 \rho}} dt_2 dt_1 \\ & = \frac{1}{2d_m(p_1)d_n(p_2)} \int_A^\infty k(u + d_m^2(p_1) + d_n^2(p_2)) \\ & \quad \times \int_{B(u)}^{C(u)} (uF(\varphi) - \sinh^2 \rho)^{-\frac{1}{2}} d\varphi du \end{aligned}$$

where

$$F(\varphi) = d_m^{-2}(p_1)d_n^{-2}(p_2)(d_m^2(p_1)\sin^2 \varphi + 2d_m(p_1)d_n(p_2)\sin \varphi \cos \varphi \cosh \rho + d_n^2(p_2)\cos^2 \varphi),$$

$$A = \frac{1}{2}(-d_m^2(p_1) + d_n^2(p_2)) + \sqrt{(d_m^2(p_1) + d_n^2(p_2))^2 + 4d_m^2(p_1)d_n^2(p_2)\sinh^2 \rho}$$

and  $B(u), C(u)$  are the limits of the range where the radical in the integral is defined and real. By using (1.3) once more this is

$$\begin{aligned} & -\frac{1}{2\pi d_m(p_1)d_n(p_2)} \int_A^\infty Q'(t + d_m^2(p_1) + d_n^2(p_2)) \\ & \quad \int_{B(t)}^{C(t)} \int_{\frac{\sinh^2 \rho}{F(\varphi)}}^t (uF(\varphi) - \sinh^2 \rho)^{-\frac{1}{2}} \frac{du}{(t-u)^{\frac{1}{2}}} d\varphi dt \end{aligned}$$

which can be calculated to give the form appearing in the Theorem 1.

## 8. Proof of Corollary 1.

To extract the corollaries from Theorem 1 we need quantitative expressions for the spectral side for two different choices of test function  $Q$ . We begin by giving a different form to the double transform  $h(r_1, r_2)$ . Recalling the transforms from Theorem 1 we can write  $h(r_1, r_2)$  directly as an integral transform of  $Q$  by

$$(8.1) \quad h(r_1, r_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{4 \sinh^2 \frac{1}{2} \xi_1 + 4 \sinh^2 \frac{1}{2} \xi_2}^{\infty} \frac{Q(t)}{\sqrt{t - 4 \sinh^2 \frac{1}{2} \xi_1 - 4 \sinh^2 \frac{1}{2} \xi_2}} e^{ir_1 \xi_1 + ir_2 \xi_2} dt d\xi_1 d\xi_2$$

since the double transform  $Q_1$  of  $k(u_1 + u_2)$  is  $L(w_1 + w_2)$  as in Theorem 1. Using  $x_i = \sinh \frac{1}{2} \xi_i$  followed by a change to polar coordinates we obtain the form

$$(8.2) \quad 2 \int_0^{\infty} Q(t) \int_0^{\frac{1}{4}t} \frac{1}{\sqrt{t-4u}} \frac{\int_0^{2\pi} (\sin \vartheta \sqrt{u} + \sqrt{u \sin^2 \vartheta + 1})^{2ir_1} (\cos \vartheta \sqrt{u} + \sqrt{u \cos^2 \vartheta + 1})^{2ir_2}}{\sqrt{(u \sin^2 \vartheta + 1)(u \cos^2 \vartheta + 1)}} d\vartheta du dt$$

which we will use to estimate  $h(r_1, r_2)$  in various cases of  $r_1$  and  $r_2$ . In the case where  $r_1 = r_2 = r_0 = -\frac{1}{2}i$  the innermost integral can be calculated as  $2\pi$ , hence

$$(8.3) \quad h(r_0, r_0) = 2\pi \int_0^{\infty} Q(t) t^{\frac{1}{2}} dt.$$

For other values let  $\rho_i$  denote the real part of  $ir_i$ , which here will be between 0 and  $\frac{1}{2}$ , and consider the inner integral as

$$\begin{aligned} \int_0^{2\pi} &\ll (u+1)^{\rho_1+\rho_2} \int_0^{2\pi} \frac{d\vartheta}{\sqrt{u^2 \sin^2 \vartheta \cos^2 \vartheta + u + 1}} \\ &\ll (u+1)^{\rho_1+\rho_2-\frac{1}{2}} \int_0^{\frac{1}{2}\pi} \frac{d\vartheta}{\sqrt{\frac{u^2}{\pi^2(u+1)} \vartheta^2 + 1}} \end{aligned}$$

which is  $O((u+1)^{\rho_1+\rho_2-\frac{1}{2}} u^{-\frac{1}{2}+\epsilon})$  for  $u \geq 1$  and  $O((u+1)^{\rho_1+\rho_2-\frac{1}{2}})$  for  $u < 1$ . If we now break the outer integrals accordingly then

$$(8.4) \quad h(r_1, r_2) \ll 1 + \int_1^{\infty} Q(t) t^{\rho_1+\rho_2-\frac{1}{2}+\epsilon} dt$$

for any  $\epsilon > 0$ , since

$$\int_1^t \frac{du}{\sqrt{u(u+1)}(t-u)} \ll t^{-\frac{1}{2}+\epsilon}.$$

To consider Corollary 1, let  $Q$  be  $C^\infty$  on  $[0, \infty)$  taking the values 1 on  $[0, X]$  and 0 on  $[X+Y, \infty)$ , and with derivatives of all orders bounded by

$$Q^{(\nu)}(t) \ll Y^{-\nu}, \quad \nu = 0, 1, 2, \dots,$$

for  $X \geq 1$  and  $Y < X$  some fractional power of  $X$  to be optimised later. With this choice of  $Q$  the transform

$$L(w) = \int_w^\infty \frac{Q(u) du}{\sqrt{u-w}}$$

is zero outside  $[0, X+Y]$ , less than or equal to  $2X^{\frac{1}{2}}$ , and has derivatives of all orders bounded by

$$L^{(\nu)}(w) = \int_w^{X+Y} \frac{Q^{(\nu)} du}{\sqrt{u-w}} \ll X^{\frac{1}{2}} Y^{-\nu}, \quad \nu = 0, 1, 2, \dots,$$

for  $w \in [0, X+Y]$ . If we now use integration by parts many times in both variables we find

$$\begin{aligned} h(r_1, r_2) &= \int_{-\infty}^\infty \int_{-\infty}^\infty L(4 \sinh^2 \frac{1}{2} \xi_1 + 4 \sinh^2 \frac{1}{2} \xi_2) e^{ir_1 \xi_1 + ir_2 \xi_2} d\xi_1 d\xi_2 \\ &= i^{\nu_1 + \nu_2} \int_{-\infty}^\infty \int_{-\infty}^\infty \frac{\partial^{\nu_1 + \nu_2}}{\partial \xi_1^{\nu_1} \partial \xi_2^{\nu_2}} L(4 \sinh^2 \frac{1}{2} \xi_1 + 4 \sinh^2 \frac{1}{2} \xi_2) \\ &\quad \times e^{ir_1 \xi_1 + ir_2 \xi_2} r_1^{-\nu_1} r_2^{-\nu_2} d\xi_1 d\xi_2 \end{aligned}$$

for any  $\nu_1, \nu_2$  since all boundary terms vanish. The derivatives have the form

$$\begin{aligned} (8.6) \quad & \frac{\partial^{\nu_1 + \nu_2}}{\partial \xi_1^{\nu_1} \partial \xi_2^{\nu_2}} L(4 \sinh^2 \frac{1}{2} \xi_1 + 4 \sinh^2 \frac{1}{2} \xi_2) \\ &= \sum_{i=1}^{\nu_1} \sum_{j=1}^{\nu_2} L^{(i+j)}(4 \sinh^2 \frac{1}{2} \xi_1 + 4 \sinh^2 \frac{1}{2} \xi_2) \\ &\quad P_{i, \nu_1}(\sinh \xi_1, \cosh \xi_1) P_{j, \nu_2}(\sinh \xi_2, \cosh \xi_2) \end{aligned}$$

where  $P_{i,\nu}$  are homogeneous polynomials of degree  $i$ , so

$$\begin{aligned} & \frac{\partial^{\nu_1+\nu_2}}{\partial \xi_1^{\nu_1} \partial \xi_2^{\nu_2}} L(4 \sinh^2 \frac{1}{2} \xi_1 + 4 \sinh^2 \frac{1}{2} \xi_2) \\ & \ll X^{\frac{1}{2}} \sum_{i=1}^{\nu_1} \sum_{j=1}^{\nu_2} Y^{-i-j} (1 + \sinh^2 \frac{1}{2} \xi_1)^i (1 + \sinh^2 \frac{1}{2} \xi_2)^j \end{aligned}$$

and

$$h(r_1, r_2) \ll X^{\frac{1}{2}} \left( \frac{X}{r_1 Y} \right)^{\nu_1} \left( \frac{X}{r_2 Y} \right)^{\nu_2}$$

for any  $\nu_1, \nu_2 > 0$ . If both  $r_1$  and  $r_2$  are larger than  $X^{1+\epsilon} Y^{-1}$  for some  $\epsilon > 0$  we now have

$$(8.7) \quad h(r_1, r_2) \ll (r_1 r_2)^{-A} X^{-B}$$

for any  $A, B > 0$ , and should just one of them, say  $r_1$ , be large we similarly obtain

$$(8.8) \quad h(r_1, r_2) \ll r_1^{-A} X^{-B}$$

uniformly in  $r_2$ , by taking  $\nu_2 = 0$ .

Consider now the spectral side of Theorem 1, which is

$$\begin{aligned} & \sum_{j,k} h(r_j, r_k) \langle |\varphi_j|^2, |\varphi_k|^2 \rangle - k(0) \text{Vol}(\Gamma \backslash \mathfrak{H}) \\ & = \sum_{j,k} h(r_j, r_k) \langle |\varphi_j|^2, |\varphi_k|^2 \rangle + O(X^{-\frac{1}{2}}) \end{aligned}$$

by

$$k(0) = -\frac{1}{\pi} \int_0^\infty Q'(u) u^{-\frac{1}{2}} du \ll X^{-\frac{1}{2}}.$$

For the more complicated remaining expression we have the general bound  $\|\varphi_j\|_\infty \ll r_j^{\frac{1}{2}}$  (see [8]), so for  $r_j, r_k \geq 1$

$$\langle |\varphi_j|^2, |\varphi_k|^2 \rangle \ll \min(r_j^{\frac{1}{2}}, r_k^{\frac{1}{2}}) \ll (r_j r_k)^{\frac{1}{4}}.$$

The contribution from terms where either  $r_j$  or  $r_k$  is greater than  $X^{1+\epsilon} Y^{-1}$  can thus be seen to be extremely small, say  $O(X^{-100})$ , by taking  $A$  and/or  $B$  sufficiently large in (8.7) and (8.8).

All other terms have  $|r_j|$  and  $|r_k|$  smaller than  $X^{1+\epsilon}Y^{-1}$ . For  $r_j = r_k = -\frac{1}{2}i$  we have

$$h(r_0, r_0)\langle |\varphi_0|^2, |\varphi_0|^2 \rangle = \frac{2\pi}{\text{Vol}(\Gamma \backslash \mathfrak{H})} \int_0^\infty Q(t)t^{\frac{1}{2}} dt = \frac{2\pi X^{\frac{3}{2}}}{\text{Vol}(\Gamma \backslash \mathfrak{H})} + O(X^{\frac{1}{2}}Y),$$

which will supply the main term in Corollary 1. The remaining terms where one or both of  $r_j, r_k$  are complex can be bounded using (1.6) and (8.4) by  $O(X^{\rho_1+1} + X^{2+\epsilon}Y^{-1})$  since the inner product is absolutely bounded. Again using (1.6) we have

$$\sum_{r_j \ll X^{1+\epsilon}Y^{-1}} r_j^{\frac{1}{4}} \ll \left(\frac{X}{Y}\right)^{\frac{9}{4}} X^\epsilon$$

so the contribution from situations when both  $r_j, r_k$  are real but small is  $O(X^{5+\epsilon}Y^{-9/2})$  again by (8.4). Thus

$$(8.9) \quad \sum_{j,k} h(r_j, r_k)\langle |\varphi_j|^2, |\varphi_k|^2 \rangle \\ = \frac{2\pi X^{\frac{3}{2}}}{\text{Vol}(\Gamma \backslash \mathfrak{H})} + O(X^{\frac{1}{2}}Y + X^{2+\epsilon}Y^{-1} + X^{\rho_1+1} + X^{5+\epsilon}Y^{-9/2})$$

which is optimised by choosing  $Y = X^{9/11}$ , giving the right-hand side of Corollary 1.

The first geometric expression from Theorem 1,

$$\sum_{P \in \mathcal{P}} \log p \sum_{(m,n) \neq (0,0)} \frac{Q(d_m^2(p) + d_n^2(p))}{(d_m^2(p) + d_n^2(p))^{\frac{1}{2}}},$$

can be estimated as being well within the error terms already obtained; discarding geodesics shorter than 1 as contributing no more than  $O(\log^2 X)$ , for larger values of  $\log p$  the sum in  $n$  and  $m$  is  $O(\log^2 X / \log^2 p)$ , and hence the whole sum is  $O(X^{1+\epsilon})$  by (1.7)

It remains to demonstrate that the geometric terms remaining have the structure shown. If we construct the geometric side of Theorem 1 for the characteristic function of  $[0, X]$  in place of  $Q$  then by positivity this is bounded above by the expression in Theorem 1 for  $Q$  as above, and bounded below by the analogous expression for  $X - Y$  in place of  $X$ . Changing  $X$  to  $X - Y$  in the above analysis does not affect the estimates obtained, so we may calculate the integrals using the characteristic function in place

of  $Q$ , claiming the same asymptotic expression (8.9). Considering the first integrals, corresponding to intersection points, we note that the integrals vanish for  $d_m^2(p_1) + d_n^2(p_2) > X$ , and otherwise are

$$\int_0^{X-a} \frac{du}{\sqrt{u(u-b)(u-c)}}$$

for

$$\begin{aligned} a &= d_m^2(p_1) + d_n^2(p_2), \\ b &= -\frac{1}{2}(d_m^2(p_1) + d_n^2(p_2)) + \frac{1}{2}\sqrt{d_m^4(p_1) + 2d_m^2(p_1)d_n^2(p_2)\cos 2\vartheta + d_n^4(p_2)}, \\ c &= -\frac{1}{2}(d_m^2(p_1) + d_n^2(p_2)) - \frac{1}{2}\sqrt{d_m^4(p_1) + 2d_m^2(p_1)d_n^2(p_2)\cos 2\vartheta + d_n^4(p_2)}. \end{aligned}$$

We may calculate this as an elliptic integral by using the standard variable change  $x = u^{\frac{1}{2}}(u-b)^{-\frac{1}{2}}$  to give

$$\frac{2^{\frac{1}{2}}(2-\kappa^2)^{\frac{1}{2}}}{\kappa(d_m^2(p_1) + d_n^2(p_2))^{\frac{1}{2}}} F\left(\tan^{-1} \sqrt{\frac{2-\kappa^2}{1-\kappa^2} \left(\frac{X}{d_m^2(p_1) + d_n^2(p_2)} - 1\right)}, \kappa\right)$$

with  $\kappa$  as defined in the introduction. Similarly the case of common perpendiculars is

$$\int_b^{X-a} \frac{du}{\sqrt{u(u-b)(u-c)}}$$

for

$$\begin{aligned} a &= d_m^2(p_1) + d_n^2(p_2), \\ b &= -\frac{1}{2}(d_m^2(p_1) + d_n^2(p_2)) + \frac{1}{2}\sqrt{d_m^4(p_1) + 2d_m^2(p_1)d_n^2(p_2)\cosh 2\rho + d_n^4(p_2)}, \\ c &= -\frac{1}{2}(d_m^2(p_1) + d_n^2(p_2)) - \frac{1}{2}\sqrt{d_m^4(p_1) + 2d_m^2(p_1)d_n^2(p_2)\cosh 2\rho + d_n^4(p_2)}, \end{aligned}$$

which vanishes for  $p_1, p_2$  and  $\kappa$  as described in Corollary 1, and in other cases can be calculated using  $x = (u-b)^{\frac{1}{2}}u^{-\frac{1}{2}}$  as

$$\frac{2(2\kappa^2-1)^{\frac{1}{2}}}{(d_m^2(p_1) + d_n^2(p_2))^{\frac{1}{2}}} F\left(\sec^{-1} \sqrt{\frac{2\kappa^2-1}{1-\kappa^2} \left(\frac{X}{d_m^2(p_1) + d_n^2(p_2)} - 1\right)}, \kappa\right).$$

This completes the proof of Corollary 1.

### 9. Proof of Corollary 2.

Choose  $Q(u)$  now to be 1 on  $[X, X + Y]$ , zero outside  $[X - Y, X + 2Y]$  and with derivatives of all orders bounded by

$$Q^{(\nu)}(u) \ll Y^{-\nu}, \quad \nu = 0, 1, 2, \dots$$

Let  $P_1, P_2$  be two primitive closed geodesics,  $\vartheta$  be an intersection angle,  $\rho$  the length of a common perpendicular, and let  $I_1(P_1, P_2, \vartheta)$ ,  $I_2(P_1, P_2, \rho)$  denote the corresponding integrals from Theorem 1. From the positivity of the terms,

$$\sum_{\vartheta} I_1(P_1, P_2, \vartheta) + \sum_{\rho} I_2(P_1, P_2, \rho)$$

is bounded above by the geometric side of Theorem 1, and hence by

$$\begin{aligned} \sum_{j,k} h(r_j, r_k) \langle |\varphi_j|^2, |\varphi_k|^2 \rangle - k(0) \text{Vol}(\Gamma \backslash \mathfrak{H}) \\ = \sum_{j,k} h(r_j, r_k) \langle |\varphi_j|^2, |\varphi_k|^2 \rangle + O(X^{-\frac{1}{2}}). \end{aligned}$$

Using analysis similar to the proof of Corollary 1, differing only in the size of  $L^{(\nu)}(w)$  which is now  $O(X^{-\frac{1}{2}}Y^{1-\nu})$ , we find that this spectral sum is

$$\frac{2\pi}{\text{Vol}(\Gamma \backslash \mathfrak{H})} \int_0^\infty Q(t) t^{\frac{1}{2}} dt + O(X^{1+\epsilon} + X^{\rho_1} Y + X^{4+\epsilon} Y^{-\frac{7}{2}}).$$

Since the leading term is at least as large as  $YX^{\frac{1}{2}}$  and no larger than  $2YX^{\frac{1}{2}}$  (up to an error already present) it dominates over the error term if  $Y = X^\alpha$  for  $\alpha > 7/9$ , and hence we may now claim that for any constant  $c > 4\pi/\text{Vol}(\Gamma \backslash \mathfrak{H})$ , for sufficiently large  $X$ ,

$$\sum_{\vartheta} I_1(P_1, P_2, \vartheta) + \sum_{\rho} I_2(P_1, P_2, \rho) \leq cX^{\alpha+\frac{1}{2}}.$$

If we choose now  $X = d^2(p_1) + d^2(p_2)$  and suppose that  $X \geq 1$  then

$$u^2 \leq u(d^2(p_1) + d^2(p_2)) + d^2(p_1)d^2(p_2) \sin^2 \vartheta$$

in the region of integration, and

$$\begin{aligned}
 I_1(P_1, P_2, \vartheta) &\geq \int_0^Y \frac{du}{\sqrt{u^3 + u^2(d^2(p_1) + d^2(p_2)) + ud^2(p_1)d^2(p_2) \sin^2 \vartheta}} \\
 &\geq 2^{-\frac{1}{2}} \int_0^Y \frac{du}{\sqrt{u^2(d^2(p_1) + d^2(p_2)) + ud^2(p_1)d^2(p_2) \sin^2 \vartheta}} \\
 &= 2^{\frac{1}{2}} \int_0^{Y^{\frac{1}{2}}} \frac{dt}{\sqrt{t^2(d^2(p_1) + d^2(p_2)) + d^2(p_1)d^2(p_2) \sin^2 \vartheta}} \\
 &= \left( \frac{2}{d^2(p_1) + d^2(p_2)} \right)^{\frac{1}{2}} \sinh^{-1} \left( \frac{Y(d^2(p_1) + d^2(p_2))}{d^2(p_1)d^2(p_2) \sin^2 \vartheta} \right)^{\frac{1}{2}}.
 \end{aligned}$$

We treat  $I_2(P_1, P_2, \rho)$  similarly, and find

$$I_2(P_1, P_2, \rho) \geq \left( \frac{2}{d^2(p_1) + d^2(p_2)} \right)^{\frac{1}{2}} \cosh^{-1} \left( \frac{Y(d^2(p_1) + d^2(p_2))}{d^2(p_1)d^2(p_2) \sinh^2 \rho} \right)^{\frac{1}{2}},$$

so combining expressions we find

$$\begin{aligned}
 \sum_{\vartheta} \sinh^{-1} \left( \frac{(d^2(p_1) + d^2(p_2))^{\alpha+1}}{d^2(p_1)d^2(p_2) \sin^2 \vartheta} \right)^{\frac{1}{2}} + \sum_{\rho} \cosh^{-1} \left( \frac{(d^2(p_1) + d^2(p_2))^{\alpha+1}}{d^2(p_1)d^2(p_2) \sinh^2 \rho} \right)^{\frac{1}{2}} \\
 \leq \frac{c}{\sqrt{2}} (d^2(p_1) + d^2(p_2))^{\alpha+1}
 \end{aligned}$$

for sufficiently large  $d^2(p_1) + d^2(p_2)$ . The first part of Corollary 2 follows now by redefining  $\alpha$  and  $c$  in the obvious way; the second part by dropping the sums over  $\vartheta$  and  $\rho$ .

## BIBLIOGRAPHY

- [1] A. BEARDON, *The Geometry of Discrete Groups*, Springer, 1983.
- [2] H. IWANIEC, Prime geodesic theorem, *J. reine. angew. Math.*, 349 (1984), 136–159.
- [3] H. IWANIEC, *Introduction to the Spectral Theory of Automorphic Forms*, Biblioteca de la Revista Matemática Iberoamericana (1995).
- [4] J. LEHNER, *Discontinuous Groups and Automorphic functions*, Amer. Math. Soc. (1964).
- [5] W. LUO and P. SARNAK, Quantum ergodicity of eigenfunctions on  $\mathrm{PSL}_2(\mathbb{Z}) \backslash \mathbb{H}$ , *IHES Publ.*, 81 (1995 207–237).

- [6] N. PITT, Talk given at the XIV Escola de Algebra, IMPA, Rio de Janeiro, Aug. 1996.
- [7] P. SARNAK, Arithmetic Quantum Chaos, Israel Mathematical Conference Proceedings, 8 (1995).
- [8] P. SARNAK, Class numbers of indefinite binary quadratic forms, J. Number Theory, 15 (1982), 229–247.
- [9] A. SEGER and C. SOGGE, Bounds for eigenfunctions of differential operators, Indiana Univ. Math. J., 38 (1989), 669–682.
- [10] A. SELBERG, Collected Papers, Vol. I, Springer, 1989.
- [11] S. ZELDITCH, Selberg Trace Formulae, Pseudodifferential operators and geodesic periods of automorphic forms, Duke Math. J. (1988), 295–344.

Manuscrit reçu le 15 octobre 1997,  
accepté le 3 juin 1998.

Nigel J. E. PITT,  
Universidade de Brasília  
Departamento de Matemática  
Brasília, DF 70910-900 (Brasil).  
pitt@mat.unb.br