Algebra & Number Theory

Volume 17
2023
No. 7

Intersecting geodesics on the modular surface

Junehyuk Jung and Naser Talebizadeh Sardari

msp

ALGEBRA AND NUMBER THEORY 17:7 (2023)
https://doi.org/10.2140/ant.2023.17.1325

Intersecting geodesics on the modular surface

Junehyuk Jung and Naser Talebizadeh Sardari

We introduce the *modular intersection kernel*, and we use it to study how geodesics intersect on the full modular surface $\mathbb{X} = \mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}$. Let C_d be the union of closed geodesics with discriminant d and let $\beta \subset \mathbb{X}$ be a compact geodesic segment. As an application of Duke's theorem to the modular intersection kernel, we prove that $\{(p, \theta_p) : p \in \beta \cap C_d\}$ becomes equidistributed with respect to $\sin\theta\, ds\, d\theta$ on $\beta \times [0, \pi]$ with a power saving rate as $d \to +\infty$. Here θ_p is the angle of intersection between β and C_d at p. This settles the main conjectures introduced by Rickards(2021).

We prove a similar result for the distribution of angles of intersections between C_{d_1} and C_{d_2} with a power-saving rate in d_1 and d_2 as $d_1 + d_2 \to \infty$. Previous works on the corresponding problem for compact surfaces do not apply to $\mathbb{X}$, because of the singular behavior of the modular intersection kernel near the cusp. We analyze the singular behavior of the modular intersection kernel by approximating it by general (not necessarily spherical) point-pair invariants on $\mathrm{PSL}_2(\mathbb{Z})\backslash \mathrm{PSL}_2(\mathbb{R})$ and then by studying their full spectral expansion.

1. Introduction

Let Y be a negatively curved surface of finite area. The prime geodesic theorem [Sarnak 1980] states that the number of primitive closed geodesics having length less than L, which we denote by $\pi(L)$, satisfies

$$\pi(L) \sim \frac{e^L}{L},$$

as $L \to \infty$. A natural problem is to understand how primitive closed geodesics of length less than L are positioned or distributed in Y as $L \to \infty$. In particular, one may ask

(1) how the number of transversal intersections $I(\alpha_1, \alpha_2)$ between two primitive closed geodesics α_1 and α_2 is distributed, or

(2) how the set of angles of intersections between α_1 and α_2 is distributed,

We thank D. Jakobson, V. Blomer, D. Milicevic, C. Pagano, M. Lee, M. Lipnowski, I. Khayutin, P. Humphries, Y. Kim, and J. Yim for valuable comments. Jung thanks A. Reid for the discussion that led to this project. Jung was supported by NSF grant DMS-1900993, and by Sloan Research Fellowship. Talebizadeh Sardari was supported partially by the National Science Foundation under Grant No. DMS-2015305, and is grateful to Max Planck Institute for Mathematics in Bonn for its hospitality and financial support.
MSC2020: 11F03, 11F70, 11F72.
Keywords: Closed geodesics, Modular forms, Intersection angles.

© 2023 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

as one varies α_1, or both α_1 and α_2? Bonahon [1986] defined the *intersection form* $i : \mathcal{C} \times \mathcal{C} \to \mathbb{R}^+$ on the space of currents $\mathcal{C}$ such that when μ_i ($i = 1, 2$) is the unique invariant measure corresponding to α_i, then $i(\mu_1, \mu_2) = I(\alpha_1, \alpha_2)$. When Y is compact, Pollicott and Sharp [2006] used an extension of the intersection form to understand the distribution of angles of self-intersections of closed geodesic α having length less than L, as $L \to \infty$. When Y is a compact hyperbolic surface, using the intersection form, Herrera Jaramillo [2015] proved that the distribution of $I(\alpha_1, \alpha_2)/(l(\alpha_1)l(\alpha_2))$ for closed geodesics α_1, α_2 of length $< L$, is concentrated near $1/(2\pi^2(g-1)) = 2/(\pi \operatorname{vol}(Y))$ with exponentially decaying tail, as $L \to \infty$. Here $l(\cdot)$ is the length function, and g is the genus of Y.

In this article, we study a refined problem:

(3) How are the locations and angles of intersections between α_1 and α_2 jointly distributed relative to α_2, as one varies α_1, or both α_1 and α_2?

Let $\mathbb{X} = \mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}$ be the full modular surface. The connection between the geometry of geodesics on $\mathbb{X}$ and number theory has a rich history. Artin [1924] discovered a relation between geometry of geodesics in $\mathbb{X}$ and continued fraction expansion. As a result, he proved that there is a hyperbolic geodesic in $\mathbb{X}$ that comes arbitrarily close to any given hyperbolic segment in $\mathbb{X}$. So this geodesic is not only dense, but dense in all directions simultaneously. Another deep connection is discovered in the spectacular work of Katok [1985]. She showed that certain holomorphic Poincaré series (introduced by Petersson) associated with closed geodesics on a Fuchsian group of the first kind, span the corresponding space of cusp forms. Moreover, she proved a formula relating the intersection angles between pairs of closed geodesics to the periods of these holomorphic Poincaré series.

On $\mathbb{X}$, primitive oriented closed geodesics are in one-to-one correspondence with conjugacy classes of primitive hyperbolic elements of $\mathrm{PSL}_2(\mathbb{Z})$. Moreover there is a bijection between the primitive hyperbolic conjugacy classes and the $\mathrm{SL}_2(\mathbb{Z})$ equivalence classes of primitive integral binary quadratic forms of nonsquare positive discriminant [Luo et al. 2009; Sarnak 1982]. So by the discriminant of a primitive closed geodesic, we mean the discriminant of the corresponding binary quadratic form. In particular, if the hyperbolic class γ is associated to the binary quadratic form Q then γ^{-1} is associated to $-Q$.

Let (x_d, y_d) be the fundamental solution of Pell's equation $x^2 - dy^2 = 4$, and let $\varepsilon_d := \frac{1}{2}(x_d + \sqrt{d}\, y_d) > 1$. Each oriented primitive closed geodesics of discriminant d has a unique lift to a closed geodesic of length $2\log\varepsilon_d$ in the unit tangent bundle $S\mathbb{X}$. Let $h(d)$ be the number of inequivalent primitive integral binary quadratic forms of discriminant d. We denote the disjoint union of these $h(d)$ closed geodesics by $\mathscr{C}_d \subset S\mathbb{X}$, which has total length $2h(d)\log\varepsilon_d$.

Note that the closed geodesic on $\mathbb{X}$ has length $\log\varepsilon_d$ or $2\log\varepsilon_d$ according as Q is or is not equivalent to $-Q$ [Duke 1988, page 75]. We now let C_d be the union of primitive (unoriented) closed geodesics of discriminant d on $\mathbb{X}$, and note that $l(C_d) = h(d)\log\varepsilon_d$ is the total length of C_d.

Theorem 1.1. *Fix $T > 100$, and let β be a compact oriented geodesic segment of length < 1 in the region determined by $y < T$ on $\mathbb{X}$. For $0 < \theta_1 < \theta_2 < \pi$, let $I_{\theta_1,\theta_2}(\beta, C_d)$ be the number of intersections between*

β and C_d with the angle between θ_1 and θ_2. (Here the angle between β and C_d at $p \in \beta \cap C_d$ is measured counterclockwise from the tangent to β at p to the tangent to C_d at p.)

Then we have

$$\frac{I_{\theta_1,\theta_2}(\beta, C_d)}{l(\beta)l(C_d)} = \frac{3}{\pi^2}\int_{\theta_1}^{\theta_2} \sin\theta d\theta + O_\epsilon\,(d^{-25/3584+\epsilon}),$$

uniformly in β, θ_1, and θ_2, under the assumption that

$$\theta_2 - \theta_1 \gg d^{-25/7168},$$

and that

$$l(\beta) \gg d^{-25/7168}.$$

(Here and elsewhere, $A \ll_\tau B$ means $|A| \le C(\tau)B$ for some constant $C(\tau)$ that depends only on τ.)

Remark 1.1. This statement is false if C_d is replaced by individual geodesics. For instance, the set of intersections between β and a closed geodesic α does not necessarily become equidistributed as $l(\alpha) \to \infty$. To see this, take a finite sheeted covering S of $\mathbb{X}$ whose genus is ≥ 2. Then according to Rivin's work [2001] there are arbitrarily long simple closed geodesics on S. Note that these simple closed geodesics must be contained in a compact part of S [Jung and Reid 2021]. This implies that there is a compact set $C \subset \mathbb{X}$ which contains arbitrarily long primitive closed geodesics. Take a geodesic segment β in $\mathbb{X} - C$. Then there are infinitely many closed geodesics which do not intersect β.

Remark 1.2. The exponent $-\frac{25}{3584}$ can be improved slightly by refining our argument (for instance, by inputting the Weyl-like subconvex bound [Petrow and Young 2019] instead of the Burgess-like subconvex bound [Heath-Brown 1980]), but in order to keep the exposition simple, we do not discuss the optimal rate in the current article.

As an immediate consequence, we deduce that the intersection points and corresponding angles become equidistributed, resolving the main conjectures introduced by Rickards [2021].

Corollary 1.2. *Fix a closed geodesic α. Then for any fixed segment $\beta \subset \alpha$, and any fixed $0 < \theta_1 < \theta_2 < \pi$, we have*

$$\lim_{d\to\infty} \frac{I_{\theta_1,\theta_2}(\beta, C_d)}{I(\alpha, C_d)} = \frac{l(\beta)}{l(\alpha)}\int_{\theta_1}^{\theta_2} \frac{\sin\theta}{2}\,d\theta.$$

Remark 1.3. Rickards's work is motivated by the work of Darmon and Vonk [2022] on the arithmetic (p-arithmetic) intersection between pairs of oriented closed geodesics on the modular surfaces (Shimura curves). The arithmetic intersection between oriented closed geodesics α_1 and α_2 of discriminants D_1 and D_2 only depends on D_1 and D_2 and the angles of intersections between α_1 and α_2. Darmon and Vonk [2022, Conjecture 2] conjectured that the p-arithmetic intersection is algebraic and belongs to the composition of the Hilbert class field of real quadratic fields of discriminants D_1 and D_2.

To prove our main results, we introduce the *modular intersection kernel*. For $\delta > 0$ and $\theta_1, \theta_2 \in (0, \pi)$, let $k_\delta^{\theta_1,\theta_2} : S\mathbb{H} \times S\mathbb{H} \to \mathbb{R}$ be the integral kernel defined by

$$k_\delta^{\theta_1,\theta_2}((x_1, \xi_1), (x_2, \xi_2)) = 1,$$

if the geodesic segments of length δ from x_i with the initial vector ξ_i intersect at an angle $\in (\theta_1, \theta_2)$, and 0 otherwise. Under the identification $S\mathbb{H} \cong \mathrm{PSL}_2(\mathbb{R})$, for a given discrete subgroup $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$, we define the *modular intersection kernel* $K_\delta^{\theta_1,\theta_2} : \Gamma\backslash \mathrm{PSL}_2(\mathbb{R}) \times \Gamma\backslash \mathrm{PSL}_2(\mathbb{R}) \to \mathbb{R}$ by taking the average of $k_\delta^{\theta_1,\theta_2}$ over Γ:

$$K_\delta^{\theta_1,\theta_2}(g_1, g_2) = \sum_{\gamma\in\Gamma} k_\delta^{\theta_1,\theta_2}(g_1, \gamma g_2).$$

The basic idea of the proof of Theorem 1.1 then is as follows. Heuristically,

$$I_{\theta_1,\theta_2}(\beta, C_d)$$

should be well approximated by

$$\frac{1}{2\delta^2} \int_{\mathscr{C}_d} \int_{\tilde{\beta}} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2, \tag{1-1}$$

where $\tilde{\beta} \subset S\mathbb{X}$ is a lift of β with either of orientations of β

$$\tilde{\beta}(t) = (\beta(t), \beta'(t)),$$

under assuming that $\beta(t)$ is parametrized by the arc length. As noted in [Luo et al. 2009], Duke's theorem [1988] can be extended to the equidistribution of $\mathscr{C}_d$ in $\mathrm{PSL}_2(\mathbb{Z})\backslash \mathrm{PSL}_2(\mathbb{R})$ as $d \to \infty$. Observing that

$$\frac{1}{2\delta^2} \int_{\tilde{\beta}} K_\delta^{\theta_1,\theta_2}(s_1, g)\, ds_1 \tag{1-2}$$

is a compactly supported function in g for compact β, (1-1) is

$$\sim \frac{l(\mathscr{C}_d)}{2\delta^2} \int_g \int_{\tilde{\beta}} K_\delta^{\theta_1,\theta_2}(s_1, g)\, ds_1\, d\mu_g,$$

which is asymptotically $(3/\pi^2) l(C_d) l(\beta) \int_{\theta_1}^{\theta_2} \sin\alpha\, d\alpha$ as $\delta \to 0$, by an explicit computation.

Note that (1-2) is a discontinuous function. Therefore, in order to obtain the rate of convergence, we need a smooth approximation of (1-2), and a quantified version of Duke's theorem with explicit dependency on the test functions. To this end, we follow the argument sketched in [Luo et al. 2009] to prove:

Theorem 1.3. *Assume that $f \in C_0^\infty(\mathrm{PSL}_2(\mathbb{Z})\backslash \mathrm{PSL}_2(\mathbb{R}))$ has support in the region determined by $y < T$. Then we have*

$$\frac{1}{l(\mathscr{C}_d)} \int_{\mathscr{C}_d} f(s)\, ds = \frac{3}{\pi^2} \int_{\mathrm{PSL}_2(\mathbb{Z})\backslash \mathrm{PSL}_2(\mathbb{R})} f(g) d\mu_g + O_\epsilon(\log T\, d^{-25/512+\epsilon} \|f\|_{W^{6,\infty}}).$$

Here $\|\cdot\|_{W^{k,p}}$ is the Sobolev norm:

$$\|f\|_{W^{k,p}} = \max_{|\alpha|\le k} \|\partial_\theta^{\alpha_1}(y\partial_x)^{\alpha_2}(y\partial_y)^{\alpha_3} f\|_{L^p}.$$

Remark 1.4. The proof of Theorem 1.1 is based on the equidistribution of the lifts of C_d in the unit tangent bundle. For this reason, one may generalize Theorem 1.1 to any surfaces and any sequence of closed geodesics whose lifts become equidistributed on the unit tangent bundle.

1A. ***Intersecting two closed geodesics.*** We now consider the number of intersections between two closed geodesics when both vary.

Theorem 1.4. *The following estimate holds uniformly in* $d_1, d_2 > 0$, *and* $0 < \theta_1 < \theta_2 < \pi$ *such that* $\theta_2 - \theta_1 \gg (d_1 d_2)^{-25/3072}$

$$\frac{I_{\theta_1,\theta_2}(C_{d_1}, C_{d_2})}{l(C_{d_1})l(C_{d_2})} = \frac{3}{\pi^2}\int_{\theta_1}^{\theta_2} \sin\theta\, d\theta + O_\epsilon((d_1 d_2)^{-25/6144+\epsilon}).$$

Note that if Γ is cocompact, then the modular intersection kernel coincides with the *intersection kernel* from [Lalley 2014] when $\theta = \pi$ and $\delta > 0$ is sufficiently small. However, when $\Gamma\backslash\mathbb{H}$ is noncompact, then they are never the same; for instance, we have $K_\delta^{\theta_1,\theta_2}(g, g) = \Omega(y)$ as $y \to \infty$ (Proposition 2.2). In particular, $K_\delta^{\theta_1,\theta_2}$ is not a Hilbert–Schmidt kernel, so the usual spectral theory does not apply. This is the main technical difficulty of dealing with the modular intersection kernel for noncompact quotients of $\mathbb{H}$. As it will be shown in the subsequent chapters, when both α_1 and α_2 are closed geodesics, $I_{\theta_1,\theta_2}(\alpha_1, \alpha_2)/(l(\alpha_1)l(\alpha_2))$ is the integral of $\delta^{-2}K_\delta^{\theta_1,\theta_2}/(l(\alpha_1)l(\alpha_2))$ over $\alpha_1 \times \alpha_2$. When α_1 and α_2 vary over closed geodesics of length $< L$, as $L \to \infty$, we expect that the integral converges to the integral of $\delta^{-2}K_\delta^{\theta_1,\theta_2}$ over $\Gamma\backslash S\mathbb{H} \times \Gamma\backslash S\mathbb{H}$, since $\alpha_1 \times \alpha_2$ becomes equidistributed in $\Gamma\backslash S\mathbb{H} \times \Gamma\backslash S\mathbb{H}$, as $L \to \infty$. However, unboundedness of the modular intersection kernel K causes issues of interchanging the limit and the integral. In particular, the argument of [Pollicott and Sharp 2006] using intersection form does not apply in this case. Hence, in order to prove Theorem 1.4, we study the full spectral expansion of $K_\delta^{\theta_1,\theta_2}(g_1, g_2)$. This is similar to the existing work on the weight m Selberg's trace formula [Hejhal 1976], except that we have to deal with all weights simultaneously, and that the modular intersection kernel is not diagonalizable in general. We go over this carefully in Section 5. Once the spectral expansion is obtained, the integral of $\delta^{-2}K_\delta^{\theta_1,\theta_2}$ over $\alpha_1 \times \alpha_2$ becomes a linear combination of the period integrals of the form

$$\int_{\alpha_1} \phi_1\, ds \times \int_{\alpha_2} \phi_2\, ds.$$

We may now use the same estimates that we use in order to prove the effective Duke's theorem to bound these, which leads to Theorem 1.4, generalizing [Pollicott and Sharp 2006] to a noncompact hyperbolic surface.

2. The modular intersection kernel

2A. ***Parametrization.*** Recall that $\mathrm{PSL}_2(\mathbb{R})$ acts transitively on $\mathbb{H}$ and on $S\mathbb{H}$ with the fractional transformations. For $g \in \mathrm{PSL}_2(\mathbb{R})$, $z \in \mathbb{H}$ and $u \in S\mathbb{H}$ we write these actions by gz and gu. We parametrize the points of $\mathbb{H}$ and $S\mathbb{H}$ with $x+iy$ and $(x+iy, \exp(i\theta))$. Let

$$\Pi((x+iy, \exp(i\theta))) := x+iy,$$

be the projection map from $S\mathbb{H}$ to $\mathbb{H}$.

Fix $z_0 = i$ and $u_0 = (i, \exp(i\pi/2))$. Let $g = naR_\theta \in \mathrm{PSL}_2(\mathbb{R})$ be the Iwasawa decomposition where

$$n = n(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad a = a(y) = \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix}, \quad \text{and} \quad R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Then we have

$$gz_0 = x+iy \quad \text{and} \quad gu_0 = \bigl(x+iy, \exp\bigl(i\bigl(\tfrac{\pi}{2}+2\theta\bigr)\bigr)\bigr).$$

For the rest of the paper, we identify $S\mathbb{H}$ with $\mathrm{PSL}_2(\mathbb{R})$ by sending $g \in \mathrm{PSL}_2(\mathbb{R})$ to gu_0. We often use the following fact in our computation without mentioning.

Proposition 2.1. *The image under $\gamma \in \mathrm{SL}_2(\mathbb{R})$ of the geodesic segment of length δ corresponding to $g = (x, \xi)$ is the geodesic segment of length δ corresponding to γg.*

We use the volume form given by $dV = (dx\,dy\,d\theta)/y^2$. The volume of $S\mathbb{X}$ is then $\pi^2/3$.

2B. ***Preliminary estimates.*** We first recall here the definition of the modular intersection kernel described in the introduction. For $\delta > 0$ and $\theta_1, \theta_2 \in (0, \pi)$, we define the integral kernel

$$k_\delta^{\theta_1,\theta_2} : S\mathbb{H} \times S\mathbb{H} \to \mathbb{R}$$

by

$$k_\delta^{\theta_1,\theta_2}((x_1, \xi_1), (x_2, \xi_2)) = 1,$$

if the geodesic segment of length δ on $\mathbb{H}$ from x_1 with the initial vector ξ_1 and the segment from x_2 with the initial vector ξ_2 intersect at an angle $\in (\theta_1, \theta_2)$, and 0 otherwise. Here the angle of the intersection of geodesic segments l_1 and l_2 at $p \in l_1 \cap l_2$ is measured counterclockwise from l_1 to l_2. Under the identification $S\mathbb{H} \cong \mathrm{PSL}_2(\mathbb{R})$ from Section 2A, we note here that

$$k_\delta^{\theta_1,\theta_2}(gg_1, gg_2) = k_\delta^{\theta_1,\theta_2}(g_1, g_2)$$

for any $g, g_1, g_2 \in \mathrm{PSL}_2(\mathbb{R})$.

Now for a given discrete subgroup $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$, we define the modular intersection kernel $K_\delta^{\theta_1,\theta_2}$: $\Gamma\backslash \mathrm{PSL}_2(\mathbb{R}) \times \Gamma\backslash \mathrm{PSL}_2(\mathbb{R}) \to \mathbb{R}$ by taking the average of $k_\delta^{\theta_1,\theta_2}$ over Γ:

$$K_\delta^{\theta_1,\theta_2}(g_1, g_2) = \sum_{\gamma\in\Gamma} k_\delta^{\theta_1,\theta_2}(g_1, \gamma g_2).$$

Note that when Γ is cocompact, and $\delta > 0$ is less than a half of the injectivity radius of $\Gamma\backslash\mathbb{H}$, we have $K_\delta^{\theta_1,\theta_2} \le 1$. However, when $\Gamma\backslash\mathbb{H}$ is noncompact, $K_\delta^{\theta_1,\theta_2}(g_1, g_2)$ becomes arbitrarily large near the diagonal $g_1 = g_2$ as $y_1, y_2 \to \infty$. This is illustrated in the following proposition when $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$.

Proposition 2.2. *Fix $0 < \theta < \pi$. Then for any $1 > \delta > 0$, we have*

$$K_\delta^{0,\theta}(g, g) = \Omega_\theta(\delta y).$$

Proof. Consider

$$g = (\mathrm{R}e^{i(\pi/2+\alpha(\delta))}, e^{i\alpha(\delta)}) \in S\mathbb{H},$$

where $\alpha(\delta)$ is chosen such that the geodesic segment

$$\beta_g := \left\{\mathrm{R}e^{i\theta} : \left|\theta - \tfrac{\pi}{2}\right| < \alpha(\delta)\right\} \subset \mathbb{H}$$

has length δ. Note that the length of the segment does not depend on R and that $\alpha(\delta) \sim \delta$ as $\delta \to 0$. From this, we infer that β_g and $\beta_g + n$ with $0 < n \ll R\delta$ intersect.

The angle of intersection is explicitly given by $2 \arcsin \frac{n}{R}$. So for all sufficiently small $0 < \delta < \theta$, we have

$$k_\delta^{\theta_1,\theta_2}\left(g, \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} g\right) = 1,$$

for $0 < n \ll R\delta$. This implies that

$$K_\delta^{\theta_1,\theta_2}(g, g) \gg \delta R \gg \delta y. \qquad \square$$

In view of Proposition 2.2, the following proposition provides a nice upper bound of the modular intersection kernel.

Proposition 2.3. *Let $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ and let $1 > \delta > 0$. Let h be a compactly supported function on $S\mathbb{H}$, where we assume that $h((\,\cdot\,, \xi))$ is supported in $B_\delta(i)$ for any $\xi \in S^1$. Define $H : \Gamma\backslash S\mathbb{H} \times \Gamma\backslash S\mathbb{H}$ by*

$$H(g_1, g_2) = \sum_{\gamma\in\Gamma} h(g_1^{-1}\gamma g_2)$$

for $g_1, g_2 \in \Gamma\backslash \mathrm{PSL}_2(\mathbb{R})$. Then for $g_i = (z_i, \xi_i)$ with $\mathrm{dist}_{\Gamma\backslash\mathbb{H}}(z_1, z_2) > 2\delta$, we have

$$H(g_1, g_2) = 0.$$

When $y_1 > 0$ and $y_2 > 0$ are sufficiently large, we have

$$H(g_1, g_2) \ll \delta\sqrt{y_1 y_2}\|h\|_{L^\infty}.$$

Proof. If $H > 0$, then there exists $\gamma \in \Gamma$ such that

$$h(g_1^{-1}\gamma g_2) > 0.$$

This implies that the balls of radius δ centered at z_1 and γz_2 intersect, hence

$$\mathrm{dist}_{\mathbb{H}}(z_1, \gamma z_2) < 2\delta,$$

which contradicts the assumption.

Now to prove the second estimate, we first note that when y_2 is sufficiently large, we have $y(\gamma g_2) < 1$ unless $\gamma = \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)$. Therefore $h(g_1^{-1}\gamma g_2) > 0$ only if $\gamma = \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)$. Note that $h(g_1^{-1}\gamma g_2) = 1$ holds only if $\mathrm{dist}_{\mathbb{H}}(z_1, n + z_2) < 2\delta$. This is equivalent to

$$\operatorname{arccosh}\left(1 + \frac{(n + x_2 - x_1)^2 + (y_1 - y_2)^2}{y_1 y_2}\right) < 2\delta,$$

and so

$$(n + x_2 - x_1)^2 < y_1 y_2(\cosh(2\delta) - 1) - (y_1 - y_2)^2 \le y_1 y_2(\cosh(2\delta) - 1),$$

from which we infer that there are at most $\ll \delta\sqrt{y_1 y_2}$ choices of γ which makes $h(g_1, \gamma g_2) > 0$. □

Now we analyze the modular intersection kernel when one variable is assumed to be contained in a compact set. We first note that if δ is less than half of the injectivity radius of g_0 in $\Gamma\backslash S\mathbb{H}$, then for each $g \in S\mathbb{H}$, there is at most one $\gamma \in \Gamma$ such that

$$k_\delta^{\theta_1,\theta_2}(g_0, \gamma g) \neq 0.$$

Therefore $K_\delta^{\theta_1,\theta_2}(g_0, \cdot)$ coincides with $k_\delta^{\theta_1,\theta_2}(g_0, \cdot)$ in the 2δ-neighborhood of g_0, which is a translation of $k_\delta^{\theta_1,\theta_2}((i, i), \cdot)$ around (i, i).

Lemma 2.4. *For* $0 < \theta_1 < \theta_2 < \pi$, *we have*

$$\int_{\mathbb{H}} k_\delta^{\theta_1,\theta_2}((i, i), g)\, dV = (\cos\theta_1 - \cos\theta_2)\delta^2.$$

Assume that $0 < \delta < 1$. *Then for any* $\varepsilon = o(\delta)$ *and* $\varepsilon = o(\theta_2 - \theta_1)$ *there exist a smooth majorant* $M_\delta^{\theta_1,\theta_2}$ *and a smooth minorant* $m_\delta^{\theta_1,\theta_2}$, *i.e.,*

$$0 \le m_\delta^{\theta_1,\theta_2} \le k_\delta^{\theta_1,\theta_2}((i, i), \cdot) \le M_\delta^{\theta_1,\theta_2},$$

such that

$$\int m_\delta^{\theta_1,\theta_2}\, dV \quad \text{and} \quad \int M_\delta^{\theta_1,\theta_2}\, dV$$

are both

$$(\cos\theta_1 - \cos\theta_2)\delta^2(1 + O(\varepsilon)),$$

and that

$$\|m_\delta^{\theta_1,\theta_2}\|_{W^{k,\infty}} + \|M_\delta^{\theta_1,\theta_2}\|_{W^{k,\infty}} = O_k(\varepsilon^{-k}).$$

Proof. Note that the action of the geodesic flow of time t on $S\mathbb{H} = \mathrm{PSL}_2(\mathbb{R})$ is the multiplication from the right by $a(e^t)$. For given $\varphi \in (\theta_1, \theta_2)$, we describe the collection of $g \in \mathrm{PSL}_2(\mathbb{R})$ for which the corresponding geodesic segment of length δ intersects $\{iy : e^\delta > y > 1\}$ transversally at angle φ. Note that this happens only when

$$g a(e^{t_2/2}) = \begin{cases} a(e^{t_1/2}) R_{\varphi/2}, \\ a(e^{t_1/2}) R_{(\varphi+\pi)/2}. \end{cases}$$

for some $0 < t_1, t_2 < \delta$. Hence

$$g = \begin{cases} a(e^{t_1/2}) R_{\varphi/2} a(e^{-t_2/2}), \\ a(e^{t_1/2}) R_{(\varphi+\pi)/2} a(e^{-t_2/2}). \end{cases}$$

Consider $\Psi : AKA \to \mathrm{PSL}_2(\mathbb{R})$ given by

$$(t_1, \varphi, t_2) \mapsto a(e^{t_1/2}) R_{\varphi/2} a(e^{-t_2/2})$$

The determinant of the Jacobian of Ψ is a nonzero multiple of $|\sin\varphi|$ (we refer the readers to the Appendix for the computation), and so this defines a local diffeomorphism away from $\varphi = 0$ and π. Observe that Ψ is injective away from $\varphi = 0$ and π. From this we infer that the support of $k_\delta^{\theta_1,\theta_2}((i,i), g)$ is the image of the open box

$$\{(t_1, \varphi, t_2) : 0 < t_1, t_2 < \delta, \theta_1 < \varphi < \theta_2 \text{ or } \theta_1 + \pi < \varphi < \theta_2 + \pi\}$$

under Ψ, and

$$\begin{aligned} \int_{\mathbb{H}} k_\delta^{\theta_1,\theta_2}((i,i), g)\, dV &= \tfrac{1}{2} \int_0^\delta \int_0^\delta \int_{\theta_1}^{\theta_2} |\sin(\varphi)|\, d\varphi\, dt_1\, dt_2 + \tfrac{1}{2} \int_0^\delta \int_0^\delta \int_{\theta_1+\pi}^{\theta_2+\pi} |\sin(\varphi)|\, d\varphi\, dt_1\, dt_2 \\ &= (\cos\theta_1 - \cos\theta_2)\delta^2, \end{aligned}$$

where we used $dV = \frac{1}{2}|\sin\varphi|\, d\varphi\, dt_1\, dt_2$ ((A-1)).

Note that the support of $k_\delta^{\theta_1,\theta_2}((i,i), \cdot\,)$ is an open set which has a piecewise smooth boundary. Therefore, under the assumption that $\varepsilon = o(\delta)$ and $\varepsilon = o(\theta_2 - \theta_1)$, there exist smooth majorant and minorant whose L^1 norms are $(\cos\theta_1 - \cos\theta_2)\delta^2(1 + O(\varepsilon))$, and whose k-th derivatives are $O_k(\varepsilon^{-k})$. □

As an immediate application, we have the following corollary.

Corollary 2.5. *Fix a compact subset $C \subset \Gamma\backslash S\mathbb{H}$, and assume that δ is less than the half of the infimum of injectivity radius of $g \in C$ in $\Gamma\backslash S\mathbb{H}$. Then for any given compact geodesic segment $\beta \subset C$, and for any given $\varepsilon > 0$ which is $o(\delta)$ and $o(\theta_2 - \theta_1)$,*

$$\int_\beta K_\delta^{\theta_1,\theta_2}(s, \cdot\,)\, ds$$

admits a smooth majorant $M_{\beta,\delta}^{\theta_1,\theta_2}$ and a smooth minorant $m_{\beta,\delta}^{\theta_1,\theta_2}$ such that

$$\|m_{\beta,\delta}^{\theta_1,\theta_2}\|_{L^1}, \|M_{\beta,\delta}^{\theta_1,\theta_2}\|_{L^1} = l(\beta)(\cos\theta_1 - \cos\theta_2)\delta^2(1 + O(\varepsilon)),$$

and that

$$\|m_{\beta,\delta}^{\theta_1,\theta_2}\|_{W^{k,\infty}} + \|M_{\beta,\delta}^{\theta_1,\theta_2 i}\|_{W^{k,\infty}} = O_k(l(\beta)\varepsilon^{-k}).$$

2C. *Intersection numbers.* In this section, we prove formulas relating the number of intersections between two geodesics to the integral of the modular intersection kernel over the two geodesics.

Lemma 2.6. *Let $\alpha_i = \{\alpha_i(t) : t \in [0, l(\alpha_i))\}$ be closed geodesics in $\Gamma\backslash\mathbb{H}$ parametrized by the arc length, and let $\widetilde{\alpha}_i = \{(\alpha_i(t), \alpha_i'(t)) : t \in [0, l(\alpha_i))\} \subset S\mathbb{H}$ be the lifts of α_i for $i = 1, 2$. Then for any $\delta > 0$,*

$$I_{\theta_1,\theta_2}(\alpha_1, \alpha_2) = \frac{1}{\delta^2} \int_{\widetilde{\alpha}_2} \int_{\widetilde{\alpha}_1} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2.$$

Remark 2.1. For each α_i, there are two choices of parametrization by the arc length, namely $\alpha_i(t)$ and $\alpha_i(-t)$, but the integral does not depend on the choice of the parametrizations.

Proof. By abuse of notations, we think of each α_i with $t \in [0, l(\alpha_i))$ a geodesic segment in $\mathbb{H}$ and accordingly $\widetilde{\alpha}_i$ a corresponding curve in $S\mathbb{H}$. For a geodesic segment $\alpha \subset \mathbb{H}$ parametrized by $t \in [a, b]$, let $[\alpha] \subset \mathbb{H}$ be the biinfinite geodesic $\{\alpha(t) : t \in \mathbb{R}\}$ that contains α. Then we express the integral as follows:

$$\begin{aligned}
\frac{1}{\delta^2} \int_{\widetilde{\alpha}_2} \int_{\widetilde{\alpha}_1} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 &= \sum_{\gamma\in\Gamma} \frac{1}{\delta^2} \int_{\gamma\widetilde{\alpha}_2} \int_{\widetilde{\alpha}_1} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 \\
&= \sum_{\gamma\in\Gamma/\Gamma_{[\alpha_2]}} \frac{1}{\delta^2} \int_{\gamma\widetilde{[\alpha_2]}} \int_{\widetilde{\alpha}_1} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 \\
&= \sum_{\gamma\in\Gamma_{[\alpha_1]}\backslash\Gamma/\Gamma_{[\alpha_2]}} \sum_{\gamma'\in\Gamma_{[\alpha_1]}} \frac{1}{\delta^2} \int_{\gamma'\gamma\widetilde{[\alpha_2]}} \int_{\widetilde{\alpha}_1} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 \\
&= \sum_{\gamma\in\Gamma_{[\alpha_1]}\backslash\Gamma/\Gamma_{[\alpha_2]}} \frac{1}{\delta^2} \int_{\gamma\widetilde{[\alpha_2]}} \int_{\widetilde{[\alpha_1]}} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2.
\end{aligned}$$

Here $\Gamma_{[\alpha_i]}$ is the stabilizer subgroup of Γ with respect to $[\alpha_i]$.

Now because two geodesics in $\mathbb{H}$ may intersect at most once, for each intersection point $p \in \alpha_1 \cap \alpha_2$ on $\Gamma\backslash\mathbb{H}$, there exists a unique $\gamma \in \Gamma/\Gamma_{[\alpha_2]}$ such that α_1 and $\gamma[\alpha_2]$ intersect at a lift of p. Also, because $[\alpha_1]$ is a disjoint union of $\gamma'\alpha_1$ with $\gamma' \in \Gamma_{[\alpha_1]}$, each $\{\gamma'\gamma : \gamma' \in \Gamma_{[\alpha_1]}\}$ contains at most one $\gamma'\gamma$ such that $\gamma'\gamma[\alpha_2]$ intersects α_1.

Therefore the intersections of α_1 and α_2 are in one-to-one correspondence with $\gamma \in \Gamma_{[\alpha_1]}\backslash\Gamma/\Gamma_{[\alpha_2]}$ such that $\gamma[\alpha_2]$ intersects $[\alpha_1]$. We complete the proof by observing that

$$\int_{\gamma\widetilde{[\alpha_2]}} \int_{\widetilde{[\alpha_1]}} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 = 1,$$

if $[\alpha_1]$ and $\gamma[\alpha_2]$ intersect at an angle $\in (\theta_1, \theta_2)$, and $= 0$ otherwise. □

Now let $\beta = \{\beta(t) : t \in [0, l(\beta))\}$ be a compact geodesic segment in $\Gamma\backslash\mathbb{H}$, and let α_2 be a closed geodesic as before. Then

$$\frac{1}{\delta^2} \int_{\widetilde{\alpha}_2} \int_{\tilde{\beta}} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2$$

does not always give $I(\beta, \alpha_2)$. Instead, it is a weighted sum over the intersections of $\beta_0 := \{\beta(t) : t \in [0, l(\beta) + \delta)\}$ and α_2. We prove the following.

Lemma 2.7. *With the same notations as above, assume that* $0 < \delta < l(\beta)$ *and that* β_0 *has no self intersection. For* $0 < \theta_1 < \theta_2 < \pi$, *let* $S(\beta_0, \alpha_2)_{\theta_1,\theta_2}$ *be the set of intersections between* β_0 *and* α_2 *where the intersection angle is* $\in (\theta_1, \theta_2)$. *Then we have*

$$\frac{1}{\delta^2}\int_{\tilde{\alpha}_2}\int_{\tilde{\beta}} K_\delta^{\theta_1,\theta_2}(s_1,s_2)\,ds_1\,ds_2 = \sum_{p\in S(\beta_0,\alpha_2)_{\theta_1,\theta_2}} \min\left\{\frac{\beta^{-1}(p)}{\delta}, 1, \frac{l(\beta)+\delta-\beta^{-1}(p)}{\delta}\right\}.$$

Proof. As in the proof of Lemma 2.6, we first have

$$\begin{aligned}\frac{1}{\delta^2}\int_{\tilde{\alpha}_2}\int_{\tilde{\beta}} K_\delta^{\theta_1,\theta_2}(s_1,s_2)\,ds_1\,ds_2 &= \sum_{\gamma\in\Gamma}\frac{1}{\delta^2}\int_{\gamma\tilde{\alpha}_2}\int_{\tilde{\beta}} k_\delta^{\theta_1,\theta_2}(s_1,s_2)\,ds_1\,ds_2\\ &= \sum_{\gamma\in\Gamma/\Gamma_{[\alpha_2]}}\frac{1}{\delta^2}\int_{\gamma\widetilde{[\alpha_2]}}\int_{\tilde{\beta}} k_\delta^{\theta_1,\theta_2}(s_1,s_2)\,ds_1\,ds_2.\end{aligned}$$

Note that because we assumed that β_0 has no self-intersection, $p \in S(\beta_0,\alpha_2)_{\theta_1,\theta_2}$ is in one-to-one correspondence with $\gamma \in \Gamma/\Gamma_{[\alpha_2]}$ such that β_0 and $\gamma\widetilde{[\alpha_2]}$ intersect at p at an angle $\in (\theta_1,\theta_2)$. We denote by γ_p the γ corresponding to p. Observe that

$$\int_{\gamma\widetilde{[\alpha_2]}}\int_{\tilde{\beta}} k_\delta^{\theta_1,\theta_2}(s_1,s_2)\,ds_1\,ds_2 = 0,$$

if $\gamma\widetilde{[\alpha_2]}\cap\beta_0 = \varnothing$. So it is sufficient to prove that

$$\frac{1}{\delta^2}\int_{\gamma_p\widetilde{[\alpha_2]}}\int_{\tilde{\beta}} k_\delta^{\theta_1,\theta_2}(s_1,s_2)\,ds_1\,ds_2 = \min\left\{\frac{\beta^{-1}(p)}{\delta}, 1, \frac{l(\beta)+\delta-\beta^{-1}(p)}{\delta}\right\}.$$

This follows by observing that

$$k_\delta^{\theta_1,\theta_2}((\beta(t_1),\beta'(t_1)),(\gamma_p\alpha_2(t_2),(\gamma_p\alpha_2)'(t_2))) = 1$$

for

$$(t_1,t_2) \in (\beta^{-1}(p)-\delta, \beta^{-1}(p)) \times (\alpha_2^{-1}(p)-\delta, \alpha_2^{-1}(p)),$$

and 0 otherwise, whereas the integral over $\tilde{\beta}$ is over the range $t_1 \in (0, l(\beta))$. □

3. Spectral theory

3A. *Spectral expansion.* We first go over the spectral decomposition of $L^2(S\mathbb{X})$. Readers may find more details on the subject in [Kubota 1973; Lang 1985]. On $G = \mathrm{PSL}_2(\mathbb{R})$, there is a differential operator of order 2 that commutes with the G action,

$$\Omega = y^2\partial_x^2 + y^2\partial_y^2 + y\partial_x\partial_\theta,$$

which is called the Casimir operator. An equivariant eigenfunction of Ω is a function $f \in C^\infty(S\mathbb{X})$ that satisfies

$$\Omega f = \lambda f$$

for some $\lambda \in \mathbb{R}$, and

$$f(gR_\theta) = e^{-im\theta} f(g) \tag{3-1}$$

for some $m \in 2\mathbb{Z}$. We say that a function has weight m if it satisfies (3-1).

Each irreducible (cuspidal) subrepresentation of the right regular representation

$$\rho_g : f(h) \mapsto f(hg)$$

on $L^2(S\mathbb{X})$ is generated by an equivariant eigenfunction of Ω.

We let $\boldsymbol{E}^+$ and $\boldsymbol{E}^-$ to be the raising and lowering operator acting on equivariant functions on $L^2(S\mathbb{X})$, which are given by [Jakobson 1994]

$$\boldsymbol{E}^+ = e^{-2i\theta}(2iy\partial_x + 2y\partial_y + i\partial_\theta) \quad \text{and} \quad \boldsymbol{E}^- = e^{2i\theta}(2iy\partial_x - 2y\partial_y + i\partial_\theta). \tag{3-2}$$

Note that $\boldsymbol{E}^+$ (resp. $\boldsymbol{E}^-$) maps a weight m eigenfunction of Ω to a weight $m+2$ (resp. $m-2$) eigenfunction of Ω.

For an even integer m let

$$\psi_{s,m}(g) = y^s e^{-im\theta}.$$

Note that $\psi_{s,m}$ is invariant under the action of the unipotent upper triangular matrices. The weight m Eisenstein series is then given by

$$E_m(g, s) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \psi_{s,m}(\gamma g),$$

where $\Gamma_\infty = \left\{ \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) : n \in \mathbb{Z} \right\}$ is the stabilizer subgroup of Γ with respect to the cusp $i\infty$. Although the right-hand side of the equation is absolutely convergent only for $\mathrm{Re}(s) > 1$, the weight m Eisenstein series has a meromorphic continuation to the entire complex plane.

Let Θ be the closure of

$$\left\{ \int_{-\infty}^{\infty} h(t) E_m\big(g, \tfrac{1}{2} + it\big)\, dt : h(t) \in C_0^\infty(\mathbb{R}), m \in 2\mathbb{Z} \right\}$$

in $L^2(S\mathbb{X})$, and let

$$L^2_{\mathrm{cusp}}(S\mathbb{X}) = \left\{ f \in L^2(S\mathbb{X}) : \int_0^1 f(n(x)g)\, dx = 0 \text{ for almost every } g \in S\mathbb{X} \right\}$$

be the space of cusp forms. Then we have the decomposition

$$L^2(S\mathbb{X}) = \langle\{1\}\rangle \oplus \Theta \oplus L^2_{\mathrm{cusp}}(S\mathbb{X}),$$

where $\langle\{1\}\rangle$ is the subspace spanned by a constant function.

We express the cuspidal subspace as a direct sum of subspaces generated by Maass forms and modular forms as in [Luo et al. 2009, (1.10)],

$$L^2_{\mathrm{cusp}}(S\mathbb{X}) = \sum_{j=1}^{\infty} W_{\pi_j^0} \bigoplus \sum_{m\geq 12}\sum_{j=1}^{d_m}(W_{\pi_j^m} \oplus W_{\pi_j^{-m}}),$$

where each $W_{\pi_j^m}$ corresponds to a G and Hecke irreducible subspace of a right regular representation on L^2_{cusp}. Here d_m is the dimension of the space of holomorphic cusp forms of weight m for $\mathrm{PSL}_2(\mathbb{Z})$. Each π_j^0 corresponds to a Maass–Hecke cusp form which we denote by ϕ_j^0. For $m > 0$, π_j^m corresponds to a holomorphic Hecke cusp form ϕ_j^m. We identify a weight m function on $\Gamma\backslash\mathbb{H}$

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

with a weight m Γ-invariant function F on $\mathrm{PSL}_2(\mathbb{R})$ via

$$F(g) = y^{m/2} f(z) e^{-im\theta}. \tag{3-3}$$

When $m \geq 0$, viewing ϕ_j^m as a function on $S\mathbb{X}$, each $W_{\pi_j^m}$ is spanned by

$$\dots,\quad (\boldsymbol{E}^-)^3\phi_j^m,\quad (\boldsymbol{E}^-)^2\phi_j^m,\quad \boldsymbol{E}^-\phi_j^m,\quad \phi_j^m,\quad \boldsymbol{E}^+\phi_j^m,\quad (\boldsymbol{E}^+)^2\phi_j^m,\quad (\boldsymbol{E}^+)^3\phi_j^m,\quad \dots$$

Note that when $m > 0$, $\boldsymbol{E}^-\phi_j^m = 0$.

For $m < 0$, we set

$$W_{\pi_j^{-m}} = \overline{W_{\pi_j^m}} = \{\bar{f} : f \in W_{\pi_j^m}\}.$$

Now let

$$U_{\pi_j^0} = W_{\pi_j^0} \quad \text{and} \quad U_{\pi_j^m} = W_{\pi_j^m} \oplus W_{\pi_j^{-m}},$$

when $m > 0$. We specify an orthonormal basis of each $U_{\pi_j^m}$ as follows.

The Maass cusp form case $m = 0$: Let $-\left(\frac{1}{4}+t_j^2\right)$ be the Laplacian eigenvalue of ϕ_j^0,[†] for some real t_j. We set $\phi_{j,0}^0 = \phi_j^0$, and define $\phi_{j,l}^0$ for $l \in 2\mathbb{Z}$ inductively by

$$\boldsymbol{E}^-\phi_{j,l}^0 = (l+1-2it_j)\phi_{j,l-2}^0 \quad \text{and} \quad \boldsymbol{E}^+\phi_{j,l}^0 = (l+1+2it_j)\phi_{j,l+2}^0. \tag{3-4}$$

The holomorphic Hecke cusp form case $m > 0$: We set $\phi_{j,m}^m = \phi_j^m$ and $\phi_{j,-m}^m = \overline{\phi_j^m}$, and define $\phi_{j,l}^m$ for $l \in 2\mathbb{Z}$ inductively by

$$\boldsymbol{E}^-\phi_{j,l}^m = (l-m)\phi_{j,l-2}^m \quad \text{and} \quad \boldsymbol{E}^+\phi_{j,l}^m = (l+m)\phi_{j,l+2}^m. \tag{3-5}$$

Finally, note that we have the following relation among the weight m Eisenstein series.

$$\boldsymbol{E}^- E_m\left(g, \tfrac{1}{2}+it\right) = (m+1-2it)E_{m-2}\left(g, \tfrac{1}{2}+it\right), \quad \text{and}$$
$$\boldsymbol{E}^+ E_m\left(g, \tfrac{1}{2}+it\right) = (m+1+2it)E_{m+2}\left(g, \tfrac{1}{2}+it\right).$$

[†] Formally, it is the eigenvalue of the Laplace–Beltrami operator on $\mathbb{X}$ that corresponds to ϕ_j^0.

With these notations, we have:

Proposition 3.1. *Let $f \in L^2(S\mathbb{X})$. Then we have*

$$f(g) = \frac{3}{\pi^2}\int_{S\mathbb{X}} f(g_1)\,dg_1 + \sum_{\substack{m\ge 0\\ 2\mid m}}\sum_{j=1}^{d_m}\sum_{\substack{l\in 2\mathbb{Z}\\ |l|\ge m}} \langle f, \phi^m_{j,l}\rangle_{S\mathbb{X}}\phi^m_{j,l}(g) + \sum_{m\in 2\mathbb{Z}} \frac{1}{4\pi}\int_{-\infty}^{\infty}\bigl\langle f, E_m\bigl(\cdot, \tfrac{1}{2}+it\bigr)\bigr\rangle_{S\mathbb{X}} E_m\bigl(g, \tfrac{1}{2}+it\bigr)\,dt,$$

where we set $d_0 = +\infty$.

4. Effective equidistribution

4A. *Invariant linear form.* Define μ_d to be the integral over discriminant d oriented closed geodesics on $S\mathbb{X}$,

$$\mu_d(F) := \int_{\mathscr{C}_d} F(s)\,ds = \sum_{\mathrm{disc}(q)=d}\int_{C(q)} F(s)\,ds.$$

where $C(q) \subset S\mathbb{X}$ is the oriented closed geodesic associated to the binary quadratic form q [Luo et al. 2009, 2.3]. Then for any $F \in U_{\pi^m_j}$, we have

$$\mu_d(F) = \mu_d(\phi^m_j)\eta^m_j(F)$$

for some linear form η^m_j on $U_{\pi^m_j}$ invariant under the diagonal action [loc. cit., Section 3.7.1], which we describe below following [loc. cit., Section 3.2]. (Note that the parameter s in [loc. cit.] is replaced by $2it$ in this article for consistency.)

The Maass cusp form case $m = 0$: Let $\phi^0_{j,l}$ be the Maass form defined by (3-4). When $4\mid l$ and $l \ge 4$, we have

$$\eta^0_j(\phi^0_{j,l}) = \eta^0_j(\phi^0_{j,-l}) = \frac{(1-2it_j)(5-2it_j)\cdots(l-3-2it_j)}{(3+2it_j)(7+2it_j)\cdots(l-1+2it_j)}, \tag{4-1}$$

and $\eta^0_j(\phi^0_{j,l})$ is identically 0 if $l \equiv 2 \pmod 4$. Note that $\{\phi^0_{j,l}\}_{l\in 2\mathbb{Z}}$ is an orthogonal basis of $U_{\pi^0_j}$, and normalized so that,

$$\|\phi^0_{j,l}\|_{L^2} = \|\phi^0_j\|_{L^2}.$$

The holomorphic Hecke cusp form case $m > 0$: Let $\phi^m_{j,l}$ be the holomorphic Hecke cusp form defined by (3-5). When $m \equiv 2 \pmod 4$, η^m_j is identically 0.

When $m \equiv 0 \pmod 4$, for $l \ge 4$ with $4\mid l$,

$$\eta^m_j(\phi^m_{j,m+l}) = \eta^m_j(\phi^m_{j,-m-l}) = \frac{1\cdot 3\cdot 5\cdots(l/2-1)}{(m+1)(m+3)\cdots(m+l/2-1)}, \tag{4-2}$$

and $\eta^m_j(\phi^m_{j,m+l})$ vanishes for $l \equiv 2 \pmod 4$.

Note that $\{\phi^m_{j,l}\}_{l\in 2\mathbb{Z},|l|\ge m}$ is an orthogonal basis of $U_{\pi^m_j}$, and normalized so that

$$\|\phi^m_{j,l}\|_{L^2} = \|\phi^m_j\|_{L^2}.$$

for $l \in 2\mathbb{Z}$, $|l| \ge m$.

Eisenstein series case: By the above identities and following [Luo et al. 2009, Section 3], we have

$$\mu_d\big(E_m\big(g, \tfrac{1}{2}+it\big)\big) = \eta(m,t)\mu_d\big(E_0\big(g, \tfrac{1}{2}+it\big)\big),$$

where for $m \geq 4$ such that $4 \mid m$,

$$\eta(m,t) = \eta(-m,t) = \frac{(1-2it)(5-2it)\cdots(2m-3-2it)}{(3+2it)(7+2it)\cdots(2m-1+2it)}, \tag{4-3}$$

and $\eta(m,t)$ is identically 0 if $m \equiv 2 \pmod 4$.

4B. *Period integrals.*

4B1. *Holomorphic cusp forms.* In this section, we give an upper bound on the period integrals of holomorphic forms. We first use the results of Shintani to relate the period integrals of holomorphic cusp forms to the Fourier coefficients of half integral holomorphic forms. We then apply the result of Kohnen and Zagier [1981] which gives an explicit version of the Waldspurger's formula for the Fourier coefficients of half integral holomorphic forms. An upper bound on these period integrals is deduced by using the subconvexity bounds on the central value of the L-functions and the Ramanujan bound on the Fourier coefficients of holomorphic modular forms.

Note that $c(d)$ is identically zero when $m \equiv 2 \pmod 4$, and so we assume that $4 \mid m$. Let $\hat{\phi}_j^m$ be a normalization of the Hecke holomorphic cusp form ϕ_j^m of weight m such that $a_1 = 1$. Let

$$c(d) := \sum_{\mathrm{disc}(q)=d} \int_{C(q)} \hat{\phi}_j^m(z) q(z,1)^{m/2-1}\, dz,$$

where $\hat{\phi}_j^m(z)$ is the associated holomorphic modular form defined on the upper half plane and the integration is on the upper half plane (3-3). By [Luo et al. 2009, (2.4) page 14], we have

$$|c(d)| = |d|^{m/4-1/2} |\mu_d(\hat{\phi}_j^m)|. \tag{4-4}$$

Let

$$\theta(z, \hat{\phi}_j^m) := \sum_{d \geq 1} c(d) e(dz).$$

By [Shintani 1975, Theorem 2], $\theta(z, \phi_j^m)$ is a Hecke holomorphic cusp form of weight $(m+1)/2$ and level $\Gamma_0(4)$. By [Luo et al. 2009, (6.2), page 37], we have the following explicit version of Rallis inner product formula

$$\langle \theta(\hat{\phi}_j^m), \theta(\hat{\phi}_j^m) \rangle = \frac{(m/2-1)!}{2^m \pi^{m/2}} L\big(\tfrac{1}{2}, \phi_j^m\big) \langle \hat{\phi}_j^m, \hat{\phi}_j^m \rangle.$$

Suppose that $d = Db^2$ with D a fundamental discriminant. By [Kohnen and Zagier 1981, Theorem 1], for D a fundamental discriminant with $D > 0$ and $4 \mid m$, we have

$$\frac{c(D)^2}{\langle \theta(\hat{\phi}_j^m), \theta(\hat{\phi}_j^m) \rangle} = \frac{(m/2-1)!}{\pi^{m/2}} D^{(m-1)/2} \frac{L(1/2, \phi_j^m \otimes \chi_D)}{\langle \hat{\phi}_j^m, \hat{\phi}_j^m \rangle},$$

which implies that

$$|c(D)| = D^{(m-1)/4} \frac{(m/2-1)!}{2^{m/2}\pi^{m/2}} \big(L\big(\tfrac{1}{2}, \phi_j^m\big) L\big(\tfrac{1}{2}, \phi_j^m \otimes \chi_D\big)\big)^{1/2}.$$

By using the Ramanujan bound on the Fourier coefficients of integral weight cusp forms and the above, we have

$$|c(d)| \ll_\epsilon b^{m-1/2+\epsilon} |c(D)| \ll_\epsilon d^{m-1/4+\epsilon} \frac{(m/2-1)!}{2^{m/2}\pi^{m/2}} \big(L\big(\tfrac{1}{2}, \phi_j^m\big) L\big(\tfrac{1}{2}, \phi_j^m \otimes \chi_D\big)\big)^{1/2},$$

and so

$$|\mu_d(\hat{\phi}_j^m)| \ll_\epsilon |d|^{1/4+\epsilon} \frac{(m/2-1)!}{2^{m/2}\pi^{m/2}} \big(L\big(\tfrac{1}{2}, \phi_j^m\big) L\big(\tfrac{1}{2}, \phi_j^m \otimes \chi_D\big)\big)^{1/2},$$

by (4-4).

We now use the convexity bound

$$L\big(\tfrac{1}{2}, \phi_j^m\big) \ll_\epsilon m^{1/2+\epsilon},$$

and the subconvexity bound [Blomer et al. 2007, Theorem 1]

$$L\big(\tfrac{1}{2}, \phi_j^m \otimes \chi_D\big) \ll_\epsilon m^{(75+12\theta)/16} D^{1/2-(1/8)(1-2\theta)+\epsilon},$$

where $\theta = \frac{7}{64}$ is the best exponent toward Ramanujan conjecture for Maass forms, to see that

$$|\mu_d(\hat{\phi}_j^m)| \ll_\epsilon d^{1/4+\epsilon} \frac{(m/2-1)!}{2^{m/2}\pi^{m/2}} m^{2.64} D^{1/4-25/512}.$$

It is well-known that

$$\langle \hat{\phi}_j^m, \hat{\phi}_j^m \rangle = \frac{\Gamma(m)}{(4\pi)^m} L(1, \mathrm{sym}^2 \phi_j^m)$$

up to a constant. Hence, by Stirling's approximation

$$|\mu_d(\phi_j^m)| \ll_\epsilon d^{1/4+\epsilon} m^{2.9} D^{1/4-25/512} \ll d^{1/2-25/512+\epsilon} m^{2.9}. \tag{4-5}$$

4B2. *Maass forms.* In this section, we give an upper bound on the period integrals of Maass forms. We first recall some results of Katok and Sarnak [1993] that generalize the work of Shintani [1975] to Maass forms and related the period integrals to the Fourier coefficients of half integral weight Maass forms. Then we use an explicit version of the Waldspurger formula [Baruch and Mao 2010] and give a nontrivial bound on these period integrals by using the subconvexity bound on the central value of the L-functions and the best bound toward Ramanujan conjecture for Maass forms.

Let ϕ_j^0 be a Hecke–Maass form with $\langle \phi_j^0, \phi_j^0 \rangle = 1$ and with the Laplacian eigenvalue $-\big(\frac{1}{4} + t_j^2\big)$. For $d > 0$, let

$$\rho(d) := \frac{1}{\sqrt{8}\pi^{1/4} d^{3/4}} \sum_{\mathrm{disc}(q)=d} \int_{C(q)} \phi_j^0 \, ds$$

be the associated period integral, and let

$$\theta((u+iv),\phi_j^0) := \sum_{d\neq 0}\rho(d)W_{\operatorname{sgn}(d)/4,it_j/2}(4\pi|d|v)e(du),$$

where $W_{\operatorname{sgn}(d)/4,it_j/2}$ is the usual Whittaker function. Here $\rho(d)$ for $d<0$ is the sum of ϕ_j^0 over the CM points with the discriminant d appropriately normalized; see [Katok and Sarnak 1993, page 197] or [Sardari 2021, Section 3.3] for a detailed discussion.

Note from [Katok and Sarnak 1993] that $\theta((u+iv),\phi_j^0)$ is a weight $\frac{1}{2}$ Hecke–Maass form with the Laplacian eigenvalue $-\bigl(\frac{1}{4}+\frac{t_j^2}{4}\bigr)$. By [Katok and Sarnak 1993, (5.6), page 224] or [Luo et al. 2009, (6.4), page 38], we have the following version of the Rallis inner product formula

$$\langle\theta(\phi_j^0),\theta(\phi_j^0)\rangle = \tfrac{3}{2}\Lambda\bigl(\tfrac{1}{2},\phi_j^0\bigr),$$

where

$$\Lambda(s,\phi_j^0)=\pi^{-s}\Gamma\left(\frac{s+it_j}{2}\right)\Gamma\left(\frac{s-it_j}{2}\right)L(s,\phi_j^0)$$

is the completed L-function.

By an explicit form of Waldspurger formula [Baruch and Mao 2010, Theorem 1.4], and the best exponent toward the Ramanujan conjecture [Lester and Radziwiłł 2020, Corollary 6.1], we have

$$\frac{\rho(d)}{\langle\theta(\phi_j^0),\theta(\phi_j^0)\rangle^{1/2}} \ll_\epsilon \frac{1}{\sqrt{|d|}}\left(\frac{L(1/2,\phi_j^0\otimes\chi_D)}{L(1,\operatorname{sym}^2\phi_j^0)}\right)^{1/2} b^{7/64+\epsilon}|t_j|^{-\operatorname{sgn}(d)/4}e^{\pi|t_j|/4},$$

where $d=Db^2$ with D a fundamental discriminant. Note from Stirling's formula that

$$\Gamma\left(\frac{1/2+it_j}{2}\right)\Gamma\left(\frac{1/2-it_j}{2}\right)\ll|t_j|^{-1/2}e^{-\pi|t_j|/2},$$

from which we infer that

$$\begin{aligned}\mu_d(\phi_j^0)&\ll d^{3/4}|\rho(d)|\\ &\ll_\epsilon d^{1/4}\bigl(\Lambda\bigl(\tfrac{1}{2},\phi_j^0\bigr)\bigr)^{1/2}\left(\frac{L(1/2,\phi_j^0\otimes\chi_D)}{L(1,\operatorname{sym}^2\phi_j^0)}\right)^{1/2} b^{7/64+\epsilon}|t_j|^{-\operatorname{sgn}(d)/4}e^{\pi|t_j|/4}\\ &\ll_\epsilon d^{1/4}\bigl(L\bigl(\tfrac{1}{2},\phi_j^0\bigr)L\bigl(\tfrac{1}{2},\phi_j^0\otimes\chi_D\bigr)\bigr)^{1/2}b^{7/64+\epsilon}|t_j|^{-((\operatorname{sgn}(d)+1)/4)+\epsilon}.\end{aligned}$$

We now use the convexity bound,

$$L\bigl(\tfrac{1}{2},\phi_j^0\bigr)\ll_\epsilon|t_j|^{1/2+\epsilon},$$

and the subconvexity bound [Blomer et al. 2007, Theorem 1],

$$L\bigl(\tfrac{1}{2},\phi_j^0\otimes\chi_D\bigr)\ll_\epsilon|t_j|^{(31+4\theta+\epsilon)/16}D^{1/2-(1-2\theta)/8+\epsilon},$$

to conclude that

$$\mu_d(\phi_j^0)\ll_\epsilon d^{1/4+\epsilon}|t_j|^{3/4}b^{7/64+\epsilon}D^{1/4-25/512}\ll d^{1/2-25/512+\epsilon}|t_j|^{3/4}. \tag{4-6}$$

4B3. *Eisenstein series.* For a nonsquare integer $d \equiv 0, 1 \pmod 4$, let $d = Db^2$ where D is a fundamental discriminant. Then we have the following explicit formula for the period integral of the Eisenstein series [Zagier 1981, page 282]:[†]

$$\mu_d(E_0(\,\cdot\,, s)) = \frac{\Gamma(s/2)^2 d^{s/2} L(s, d)}{\Gamma(s)\zeta(2s)}, \tag{4-7}$$

where

$$L(s, d) = L(s, \chi_D)\biggl(\sum_{a \mid b} \mu(a)\bigl(\tfrac{D}{a}\bigr) a^{-s} \sigma_{1-2s}\bigl(\tfrac{b}{a}\bigr)\biggr). \tag{4-8}$$

Here $L(s, \chi_D)$ is the Dirichlet L-function attached to the quadratic Dirichlet character $\chi_D(\,\cdot\,) = \bigl(\frac{D}{\cdot}\bigr)$, $\mu(\,\cdot\,)$ is the Möbius function, and $\sigma_v(\,\cdot\,) = \sum_{a \mid \cdot} a^v$ is the divisor function.

Now assume that $s = \frac{1}{2} + it$ for some $t \in \mathbb{R}$. By Stirling's formula, we have

$$\frac{\Gamma(s/2)^2}{\Gamma(s)} \ll |t|^{-1/2}.$$

By the zero free region of $\zeta(2s)$ around $2s = 1 + 2it$, we have

$$|\zeta(2s)| \gg_\epsilon t^{-\epsilon}.$$

We also have the convexity bound

$$\zeta(s) \ll |t|^{1/4},$$

and we know from [Heath-Brown 1980] that

$$L\bigl(\tfrac{1}{2} + it, \chi_D\bigr) \ll_\epsilon ((|t| + 1)D)^{3/16+\epsilon}.$$

Finally, observe that we have

$$\sum_{a \mid b} \mu(a)\bigl(\tfrac{D}{a}\bigr) a^{-s} \sigma_{1-2s}\bigl(\tfrac{b}{a}\bigr) \ll_\epsilon d^\epsilon.$$

Combining all these estimates, we deduce the following estimate from (4-7) for $s = \frac{1}{2} + it$:

$$\mu_d(E_0(\,\cdot\,, s)) \ll_\epsilon d^{1/2-1/16+\epsilon}. \tag{4-9}$$

4C. ***Proof of Theorem 1.3.*** For any compactly supported smooth function $F \in C_0^\infty(S\mathbb{X})$, recall from Proposition 3.1 that we have

$$F(g) = \frac{3}{\pi^2}\int_{S\mathbb{X}} F(g_1)\,dg_1 + \sum_{\substack{m \geq 0 \\ 2 \mid m}} \sum_{j=1}^{d_m} \sum_{\substack{l \in 2\mathbb{Z} \\ |l| \geq m}} \langle F, \phi_{j,l}^m \rangle_{S\mathbb{X}} \phi_{j,l}^m(g) + \sum_{m \in 2\mathbb{Z}} \frac{1}{4\pi} \int_{-\infty}^{\infty} \bigl\langle F, E_m\bigl(\,\cdot\,, \tfrac{1}{2}+it\bigr)\bigr\rangle_{S\mathbb{X}} E_m\bigl(g, \tfrac{1}{2}+it\bigr)\,dt,$$

[†]When $b = 1$, this is a classical result due to Hecke [Siegel 1965, page 88].

and so from the discussion of Section 4A, we have

$$\mu_d(F)=\mu_d\left(\frac{3}{\pi^2}\right)\int_{S\mathbb{X}}F(g)\,dg+\sum_{\substack{m\geq 0\\ 4\,|\,m}}\sum_{j=1}^{d_m}\mu_d(\phi_j^m)\sum_{\substack{l\in 4\mathbb{Z}\\ |l|\geq m}}\langle F,\phi_{j,l}^m\rangle_{S\mathbb{X}}\eta_j^m(\phi_{j,l}^m)$$
$$+\sum_{m\in 4\mathbb{Z}}\frac{1}{4\pi}\int_{-\infty}^{\infty}\big\langle F,E_m\big(\cdot,\tfrac12+it\big)\big\rangle_{S\mathbb{X}}\eta\big(m,\tfrac12+it\big)\mu_d\big(E_0\big(\cdot,\tfrac12+it\big)\big)\,dt.$$

Firstly, we have from (4-1), (4-2), and (4-3) that $\eta_j^m(\phi_{j,l}^m)$ and $\eta\big(m,\frac12+it\big)$ are both $O(1)$. Note by successive integration by parts and Cauchy–Schwarz inequality, we have for all $N\geq 1$,

$$\langle F,\phi_{j,l}^m\rangle\ll_N(|l|^2+1)^{-N}\|F\|_{W^{2N,2}(S\mathbb{X})},$$

when $m>0$, and

$$\langle F,\phi_{j,l}^0\rangle\ll_N(|l|^2+|t_j|^2+1)^{-N}\|F\|_{W^{2N,2}(S\mathbb{X})}.$$

Likewise, assuming that the support of F is contained in $y<T$, we have

$$\big\langle F,E_m\big(\cdot,\tfrac12+it\big)\big\rangle_{S\mathbb{X}}\ll_N(|m|^2+t^2+1)^{-N}\|F\|_{W^{2N,2}(S\mathbb{X})}\log T,$$

where we used [Kubota 1973, (6.1.6)] and [Jakobson 1994, (1.6), (1.7)].

Now for $m>0$, we take $N=3$ and apply (4-5) to see that

$$\sum_{\substack{m>0\\ 4\,|\,m}}\sum_{j=1}^{d_m}\mu_d(\phi_j^m)\sum_{\substack{l\in 4\mathbb{Z}\\ |l|\geq m}}\langle F,\phi_{j,l}^m\rangle_{S\mathbb{X}}\eta_j^m(\phi_{j,l}^m)\ll_\epsilon d^{1/2-25/512+\epsilon}\|F\|_{W^{6,2}(S\mathbb{X})},$$

and for $m=0$, we take $N=2$ and apply (4-6) to deduce

$$\sum_{j=1}^{\infty}\mu_d(\phi_j^0)\sum_{l\in 4\mathbb{Z}}\langle F,\phi_{j,l}^0\rangle_{S\mathbb{X}}\eta_j^0(\phi_{j,l}^0)\ll_\epsilon d^{1/2-25/512+\epsilon}\|F\|_{W^{4,2}(S\mathbb{X})}.$$

For the Eisenstein series contribution, we take $N=2$ and apply (4-9) to see

$$\sum_{m\in 4\mathbb{Z}}\frac{1}{4\pi}\int_{-\infty}^{\infty}\big\langle F,E_m\big(\cdot,\tfrac12+it\big)\big\rangle_{S\mathbb{X}}\eta\big(m,\tfrac12+it\big)\mu_d\big(E_0\big(\cdot,\tfrac12+it\big)\big)dt\ll_\epsilon \log T\,d^{7/16+\epsilon}\|F\|_{W^{4,2}(S\mathbb{X})}.$$

Therefore Theorem 1.3 will follow once we establish the following lower bound for the total length of $\mathscr{C}_d$:

$$l(\mathscr{C}_d)=2h(d)\log\epsilon_d\gg_\epsilon d^{1/2-\epsilon}.\tag{4-10}$$

To see this, let $d=Db^2$ where D is a fundamental discriminant. Then by Dirichlet class number formula [Davenport 1967, page 50] for binary quadratic forms discriminant d (or by letting $s\to 1$ in (4-7)), we have

$$h(d)\log(\epsilon_d)=d^{1/2}L(1,d)$$

with the same $L(\cdot, d)$ given in (4-8), i.e.,

$$L(1,d)=L(1,\chi_D)\biggl(\sum_{a\,|\,b}\mu(a)\bigl(\tfrac{D}{a}\bigr)a^{-1}\sigma_{-1}\bigl(\tfrac{b}{a}\bigr)\biggr).$$

Note that

$$\sum_{a\,|\,b}\mu(a)\bigl(\tfrac{D}{a}\bigr)a^{-1}\sigma_{-1}\bigl(\tfrac{b}{a}\bigr)=\sum_{ca\,|\,b}\mu(a)\bigl(\tfrac{D}{a}\bigr)\tfrac{c}{b}=\tfrac{1}{b}\sum_{e\,|\,b}e\prod_{p\,|\,e}\bigl(1-\bigl(\tfrac{D}{p}\bigr)p^{-1}\bigr),$$

where $e=ac$, and that

$$\tfrac{1}{b}\sum_{e\,|\,b}e\prod_{p\,|\,e}\bigl(1-\bigl(\tfrac{D}{p}\bigr)p^{-1}\bigr)\gg b^{-\epsilon}.$$

Now (4-10) follows by using Siegel's lower bound [1935]

$$L(1,\chi_D)\gg_\epsilon D^{-\epsilon},$$

and this completes the proof of Theorem 1.3.

4D. ***Proof of Theorem 1.1.*** We are now ready to prove Theorem 1.1. Assume that $\beta:[0,l(\beta)]\to\mathbb{X}$ is a sufficiently short compact geodesic segment in the region determined by $y<T$ such that $\beta([-l(\beta),2l(\beta)])$ has no self intersection. (We fix T for simplicity, but it is possible to vary T with d.) For $\delta=d^{-a}$ with $a>0$ to be chosen later, such that $l(\beta)\gg\delta$, let

$$\beta_1:=\{\beta(t):t\in[0,l(\beta)-\delta]\}\quad\text{and}\quad\beta_2:=\{\beta(t):t\in[-\delta,l(\beta)]\}.$$

Then from Lemma 2.7, we have

$$\frac{1}{\delta^2}\int_{\widetilde{\alpha}_2}\int_{\widetilde{\beta}_1}K_\delta^{\theta_1,\theta_2}(s_1,s_2)\,ds_1\,ds_2\le I^{\theta_1,\theta_2}(\beta,\alpha_2)\le\frac{1}{\delta^2}\int_{\widetilde{\alpha}_2}\int_{\widetilde{\beta}_1}K_\delta^{\theta_1,\theta_2}(s_1,s_2)\,ds_1\,ds_2$$

for any closed geodesic α_2. Now define $f_1,f_2\in C_0^\infty(S\mathbb{X})$ using Lemma 2.4 by

$$f_1(g)=\frac{1}{\delta^2}\int_{\widetilde{\beta}_1}m_\delta^{\theta_1,\theta_2}(s_1^{-1}g)\,ds_1\quad\text{and}\quad f_2(g)=\frac{1}{\delta^2}\int_{\widetilde{\beta}_2}M_\delta^{\theta_1,\theta_2}(s_1^{-1}g)\,ds_1,$$

with $\varepsilon=d^{-2a}$, where we assume that $\theta_2-\theta_1\gg d^{-a}$. Note that $m(g_1^{-1}g_2)$ and $M(g_1^{-1}g_2)$ are minorant and majorant of $K_\delta^{\theta_1,\theta_2}(g_1,g_2)$ for $g_1\in\beta_i$, $g_2\in S\mathbb{X}$ for all sufficiently large d. Hence, for all sufficiently large d (independent of α_2), we have

$$\int_{\widetilde{\alpha}_2}f_1(s)\,ds\le I^{\theta_1,\theta_2}(\beta,\alpha_2)\le\int_{\widetilde{\alpha}_2}f_2(s)\,ds,$$

and so

$$\int_{\mathscr{C}_d}f_1(s)\,ds\le 2I^{\theta_1,\theta_2}(\beta,C_d)\le\int_{\mathscr{C}_d}f_2(s)\,ds,\tag{4-11}$$

where the factor 2 amounts to the fact that $\mathscr{C}_d$ is a double cover of C_d.

We now apply Theorem 1.3 to see that

$$\frac{1}{l(\mathscr{C}_d)}\int_{\mathscr{C}_d} f_i(s)ds = \frac{3}{\pi^2}\int_{S\mathbb{X}} f_i(g)d\mu_g + O_\epsilon(d^{-25/512+\epsilon}\|f_i\|_{W^{6,\infty}}).$$

Because of the choice of f_1 and f_2, we have

$$\|f_i\|_{W^{6,\infty}} \ll \varepsilon^{-6} l(\beta) \ll d^{12a} l(\beta),$$

and

$$\int_{S\mathbb{X}} f_i(g)\,d\mu_g = (\cos\theta_1 - \cos\theta_2)(l(\beta)+O(\delta))(1+O(\varepsilon)) = (\cos\theta_1-\cos\theta_2)l(\beta)(1+O(d^{-2a}))$$

by Lemma 2.4. Now we complete the proof of Theorem 1.1 for sufficiently short geodesic segments by choosing $a = \frac{25}{7168}$ and applying these estimates to (4-11). This then implies Theorem 1.1 for any geodesic segment of length < 1 by dividing the segment into finitely many sufficiently short geodesic segments, and then applying Theorem 1.1 to each of them.

5. Selberg's pretrace Formula for $\mathrm{PSL}_2(\mathbb{R})$

Let $k \in C_0^\infty(\mathrm{PSL}_2(\mathbb{R}))$, and let K be the integral kernel on $S\mathbb{X}$ defined by

$$K(g_1, g_2) = \sum_{\gamma\in\Gamma} k(g_1, \gamma g_2),$$

where $k(g_1, g_2) = k(g_1^{-1}g_2)$. The corresponding integral operator T_K acts on $f \in L^2(S\mathbb{X})$ by

$$T_K(f) := \int_{S\mathbb{X}} K(g_1, g_2) f(g_2)\,dg_2 = \int_{\mathrm{PSL}_2(\mathbb{R})} k(g_1^{-1}g_2) f(g_2)\,dg_2.$$

It follows that $T_K(f) \in L^2(S\mathbb{X})$. In this section, we study the spectral expansion of K in terms of the equivariant eigenfunctions of the Casimir operator, which are explicitly described in Section 3A. In other words, we derive Selberg's pretrace formula for $\mathrm{PSL}_2(\mathbb{Z})\backslash \mathrm{PSL}_2(\mathbb{R})$.

5A. ***Cuspidal spectrum.*** In this section, we describe explicitly the spectrum of T_K acting on the cuspidal subspace $L^2_{\mathrm{cusp}}(S\mathbb{X})$. Let $R_g(f)(x) = f(xg)$ be the right regular action of $\mathrm{PSL}_2(\mathbb{R})$ on

$$L^2_{\mathrm{cusp}}(\Gamma\backslash \mathrm{PSL}_2(\mathbb{R})) = L^2_{\mathrm{cusp}} : S\mathbb{X}).$$

Lemma 5.1. *Let π be an irreducible unitary representation of* $\mathrm{PSL}_2(\mathbb{R})$. *Then for any $f \in W_\pi \subset L^2_{\mathrm{cusp}}(S\mathbb{X})$, we have*

$$T_K(f) \in W_\pi.$$

Proof. Observe that

$$T_K(f)(g_1) = \int_{\mathrm{PSL}_2(\mathbb{R})} k(g_1, g_2) f(g_2)\,dg_2 = \int_{\mathrm{PSL}_2(\mathbb{R})} k(g_1^{-1}g_2) f(g_2)\,dg_2 = \int_{\mathrm{PSL}_2(\mathbb{R})} k(u) f(g_1 u)\,du,$$

where $u = g_1^{-1} g_2$. Hence, we have

$$T_K(f) = \int_{\mathrm{PSL}_2(\mathbb{R})} k(u) R_u(f)\, du,$$

and because $R_u(f) \in W_\pi$ for every u, we conclude that $T_K(f) \in W_\pi$. □

From 5A, for an abstract irreducible unitary representation π of $\mathrm{PSL}_2(\mathbb{R})$ and $f \in W_\pi$, we define the action of k on f by

$$k * f = \int_{\mathrm{PSL}_2(\mathbb{R})} k(u) \pi(u)(f)\, du,$$

which agrees with $T_K(f)$ when W_π is a subspace of $L^2_{\mathrm{cusp}}(S\mathbb{X})$.

Let $\psi : W_\pi \to W_{\pi'}$ be an isomorphism of representations π and π'. Note that for $f \in W_\pi$ and $f' \in W_{\pi'}$ with $\psi(f) = f'$, we have $\psi(k * f) = k * f'$. We denote by $\phi_m \in W_\pi$ the unique (up to a unit scalar) vector of norm 1 and weight m. We fix the unit scalar except for the spherical or the lowest weight vector, by using the normalized lowering and raising operator that we introduced in (3-4) and (3-5).

Now let

$$h(k, m, n, \pi) := \langle k * \phi_m, \phi_n \rangle, \tag{5-1}$$

and let $M_\pi(m, n)(g) = \langle \pi(g)\phi_m, \phi_n \rangle$ be the matrix coefficient of π. We note that $h(k, m, n, \pi)$ and $M_\pi(m, n)(g)$ do not depend on the choice of the unit scalar of the spherical or the lowest weight vector.

We recall some properties of $M_\pi(m, n)(g)$ in the following lemma.

Lemma 5.2. *We have for every $g \in \mathrm{PSL}_2(\mathbb{R})$,*

$$|M_\pi(m, n)(g)| \le 1,$$

and

$$M_\pi(m, n)(R_{\theta'} g R_\theta) = e^{-im\theta} e^{-in\theta'} M_\pi(m, n)(g).$$

Proof. We have

$$1 = |\pi(g)\phi_m|^2 = \sum_n \langle \pi(g)\phi_m, \phi_n \rangle^2,$$

from which it is immediate that $|M_\pi(m, n)(g)| \le 1$. For the second identity, we have

$$M_\pi(m, n)(R_{\theta'} g R_\theta) = \langle \pi(g)\pi(R_\theta)\phi_m, \pi(R_{-\theta'})\phi_n \rangle = e^{-im\theta} e^{-in\theta'} M_\pi(m, n)(g). \qquad □$$

Define $k_{m,n} \in C_0^\infty(\mathrm{PSL}_2(\mathbb{R}))$ by

$$k_{m,n}(g) := \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} k(R_{\theta'} g R_\theta) e^{-in\theta' - im\theta}\, d\theta'\, d\theta. \tag{5-2}$$

Note that

$$k_{m,n}(R_{\theta_1} g R_{\theta_2}) = e^{in\theta_1} k_{m,n}(g) e^{im\theta_2}. \tag{5-3}$$

The following lemma holds for every unitary irreducible representation of $\mathrm{PSL}_2(\mathbb{R})$.

Lemma 5.3. *We have*

$$h(k,m,n,\pi)=\int_{\mathrm{PSL}_2(\mathbb{R})} k_{m,n}(u)M_\pi(m,n)(u)\,du,$$

and for all nonnegative integers N_1, N_2, we have the following estimate

$$h(k,m,n,\pi)\ll_{N=N_1+N_2}(1+|m|)^{-N_1}(1+|n|)^{-N_2}\|k\|_{W^{N,1}}.$$

Proof. Recall from the definition that

$$h(k,m,n,\pi)=\int_{\mathrm{PSL}_2(\mathbb{R})} k(u)\langle\pi(u)\phi_m,\phi_n\rangle\,du=\int_{\mathrm{PSL}_2(\mathbb{R})} k(u)M_\pi(m,n)(u)\,du,$$

and so

$$\begin{aligned}h(k,m,n,\pi)&=\int_{\mathrm{PSL}_2(\mathbb{R})} k(u)M_\pi(m,n)(u)\,du.\\&=\frac{1}{4\pi^2}\int_{\mathrm{PSL}_2(\mathbb{R})}\int_\theta\int_{\theta'} k(R_{\theta'}uR_\theta)M_\pi(m,n)(R_{\theta'}uR_\theta)\,d\theta\,d\theta'\,du\\&=\frac{1}{4\pi^2}\int_{\mathrm{PSL}_2(\mathbb{R})} M_\pi(m,n)(u)\int_\theta\int_{\theta'} k(R_{\theta'}uR_\theta)e^{-im\theta}e^{-in\theta'}\,d\theta\,d\theta'\,du\\&=\int_{\mathrm{PSL}_2(\mathbb{R})} k_{m,n}(u)M_\pi(m,n)(u)\,du.\end{aligned}$$

Therefore, by integration by parts, we have

$$\begin{aligned}h(k,m,n,\pi)&\le\int_{\mathrm{PSL}_2(\mathbb{R})}|k_{m,n}(u)|\,du\\&=\int_{\mathrm{PSL}_2(\mathbb{R})}\left|\frac{1}{4\pi^2}\int_\theta\int_{\theta'} k(R_{\theta'}uR_\theta)e^{-im\theta}e^{-in\theta'}\,d\theta\,d\theta'\right|du\\&\ll_N(1+|m|)^{-N_1}(1+|n|)^{-N_2}\|k\|_{W^{N,1}},\end{aligned}$$

where we used $|M_\pi(m,n)(u)|\le 1$ from Lemma 5.2. This completes the proof of our lemma. □

5A1. *Principal series representation of* $\mathrm{SL}_2(\mathbb{R})$. For our application in the subsequent chapters, we need a refined estimate for $h(k,m,n,\pi)$ when π is a unitary principal series representation. We first give an explicit representation of $h(k,m,n,\pi)$.

Lemma 5.4. *Let W_π be a unitary principal series representation of* $\mathrm{SL}_2(\mathbb{R})$ *with the parameter* $\frac{1}{2}+it$ [Knapp 2001, Chapter VII]. *Let*

$$h(k,m,n,t):=\int_{\mathrm{PSL}_2(\mathbb{R})} k_{m,n}(g)y^{1/2+it}e^{-im\theta}\,dg, \tag{5-4}$$

where $g=na(y)R_\theta$. Then we have

$$h(k,m,n,\pi)=h(k,m,n,t).$$

Proof. We note that principal series representations are induced from the unitary characters of the upper triangular matrices to $\mathrm{PSL}_2(\mathbb{R})$ [Knapp 2001, Chapter VII]. In this model, a dense subspace of a representation is given by

$$\{f : \mathrm{PSL}_2(\mathbb{R}) \to \mathbb{C} \text{ continuous} : f(xan) = e^{(it+1/2)\log(a)} f(x)\}$$

with the norm

$$|f|^2 = \frac{1}{2\pi}\int_\theta |f(R_\theta)|^2\, d\theta,$$

and the $\mathrm{PSL}_2(\mathbb{R})$ action is given by

$$\pi(g)f(x) = f(g^{-1}x).$$

The weight m unit vectors are explicitly given by

$$\phi_m(R_\theta a(y)n) = e^{im\theta} y^{-(1/2+it)}.$$

Note that the orthonormal basis $\{\phi_m\}$ is normalized as our convention in (3-4), i.e.,

$$\boldsymbol{E}^-\phi_m = (m+1-2it)\phi_{m-2} \quad \text{and} \quad \boldsymbol{E}^+\phi_m = (m+1+2it)\phi_{m+2}.$$

With these, we first see that

$$\begin{aligned} k*\phi_m(R_{\theta'}) &= \int_{\mathrm{PSL}_2(\mathbb{R})} k(u)y(u^{-1}R_{\theta'})^{-(1/2+it)} e^{im\theta(u^{-1}R_{\theta'})}\, du \\ &= \int_{\mathrm{PSL}_2(\mathbb{R})} k(R_{\theta'}v^{-1})y(v)^{-(1/2+it)} e^{im\theta(v)}\, dv, \end{aligned}$$

where $v = u^{-1}R_{\theta'}$ and $v = R_{\theta(v)}a(y(v))n(v)$. We therefore have

$$\begin{aligned} h(k,m,n,\pi) &= \langle k*f_m, f_n\rangle \\ &= \frac{1}{2\pi}\int_{\theta'} k*f_m(R_{\theta'})\bar{f}_n(R_{\theta'})\, d\theta' \\ &= \frac{1}{2\pi}\int_{\theta'} e^{-in\theta'} \int_{\mathrm{PSL}_2(\mathbb{R})} k(R_{\theta'}v^{-1})y(v)^{-(1/2+it)}e^{im\theta(v)}\, dv\, d\theta' \\ &= \frac{1}{2\pi}\int_{\mathrm{PSL}_2(\mathbb{R})} y^{1/2+it}\int_{\theta'} e^{-in\theta'}e^{-im\theta}k(R_{\theta'}w)\, d\theta'\, dw \\ &= \int_{\mathrm{PSL}_2(\mathbb{R})} k_{m,n}(w)y^{1/2+it}e^{-im\theta}\, dw, \end{aligned}$$

where $w = v^{-1}$ and $w = na(y)R_\theta$. Note that $y = y(v)^{-1}$ and $\theta = -\theta(v)$. □

We now prove that $h(k,m,n,t)$ decays fast in all parameters uniformly.

Lemma 5.5. *Suppose that k is supported inside the compact subset $C \subset \mathrm{SL}_2(\mathbb{R})$. Then we have*

$$\int_{\mathrm{PSL}_2(\mathbb{R})} k_{m,n}(g)y^{1/2+it}e^{-im\theta}\, dg \ll_{N,C} (1+|m|)^{-N_1}(1+|n|)^{-N_2}(1+|t|)^{-N_3}\|k\|_{W^{N,\infty}}$$

for any $N_1, N_2, N_3 \geq 0$, where $N = N_1+N_2+N_3$.

Proof. From the definition, we have

$$\int_{\mathrm{PSL}_2(\mathbb{R})} k_{m,n}(g) y^{1/2+it} e^{-im\theta}\, dg = \frac{1}{4\pi} \int_{\mathbb{H}} \int_0^{2\pi} \int_0^{2\pi} k(R_{\theta_1'} n(x) a(y) R_{\theta_2'}) y^{1/2+it} e^{-in\theta_1' - im\theta_2'}\, d\theta_1'\, d\theta_2' \frac{dx\, dy}{y^2},$$

and so the statement follows from integration by parts. □

5B. *Continuous spectrum.* For $k_{m,n}$ given by (5-2), let

$$K_{m,n}(g_1, g_2) := \sum_{\gamma \in \Gamma} k_{m,n}(g_1^{-1} \gamma g_2). \tag{5-5}$$

Then we infer from (5-3) that

$$K_{m,n}(g_1 R_{\theta_1}, g_2 R_{\theta_2}) = e^{-in\theta_1} K_{m,n}(g_1, g_2) e^{im\theta_2},$$

and so it defines an integral operator that maps weight m forms to weight n forms. Denote by $S^m \subset L^2(\Gamma \backslash \mathrm{PSL}_2(\mathbb{R}))$ the space of weight m forms and by S^m_{cusp} the space of weight m forms in $L^2_{\mathrm{cusp}}(S\mathbb{X})$. We first recall the following result regarding the decomposition of $K_{m,m}$.

Theorem 5.6 [Hejhal 1976]. *The integral kernel*

$$K_{m,m}(g_1, g_2) - \frac{1}{4\pi} \int_{-\infty}^{\infty} h(k, m, m, t) E_m\big(g_1, \tfrac{1}{2} + it\big) \overline{E_m\big(g_2, \tfrac{1}{2} + it\big)}\, dt$$

defines a compact operator $S^m_{\mathrm{cusp}} \to S^m_{\mathrm{cusp}}$ that acts trivially on Θ. (Here $h(k, m, m, t)$ is given by (5-4).)

We define $\boldsymbol{E}^a$ to be $(\boldsymbol{E}^+)^a$ if $a > 0$, and $(\boldsymbol{E}^-)^{|a|}$ if $a < 0$. We have

$$\overline{\boldsymbol{E}^a} = (-\boldsymbol{E})^{-a},$$

which follows directly from (3-2). Let $c_{m,n}$ be given by

$$\boldsymbol{E}^{n-m} E_m(g, s) = c_{m,n}(s) E_n(g, s).$$

Observe that

$$\boldsymbol{E}^{n-m} y^s e^{-im\theta} = c_{m,n}(s) y^s e^{-in\theta},$$

and that

$$\overline{c_{m,n}\big(\tfrac{1}{2} + it\big)} = c_{n,m}\big(\tfrac{1}{2} + it\big) \tag{5-6}$$

for $t \in \mathbb{R}$.

Theorem 5.7. *For $m, n \in 2\mathbb{Z}$,*

$$K_{m,n}(g_1, g_2) - \frac{1}{4\pi} \int_{-\infty}^{\infty} h(k, m, n, t) E_n\big(g_1, \tfrac{1}{2} + it\big) \overline{E_m\big(g_2, \tfrac{1}{2} + it\big)}\, dt$$

defines a compact operator $S^m_{\mathrm{cusp}} \to S^n_{\mathrm{cusp}}$ that acts trivially on Θ.

Proof. Note that

$$\int \boldsymbol{E}_{g_2}^{m-n}(K(g_1,g_2)f(g_2))\,dg_2=0$$

for every g_1, $m\neq n$, and $f\in C_0^\infty(\Gamma\backslash \mathrm{PSL}_2(\mathbb{R}))$. Hence

$$T_K\boldsymbol{E}^{m-n}:C_0^\infty(\Gamma\backslash \mathrm{PSL}_2(\mathbb{R}))\to C_0^\infty(\Gamma\backslash \mathrm{PSL}_2(\mathbb{R}))$$

is an integral operator with the integral kernel

$$K'(g_1,g_2)=\sum_{\gamma\in\Gamma}k'(g_1^{-1}\gamma g_2),$$

where

$$k'(g)=(-\boldsymbol{E})^{m-n}k(g)=\overline{\boldsymbol{E}^{n-m}}k(g).$$

Then by Theorem 5.6, we see that

$$K''(g_1,g_2)=K'_{n,n}(g_1,g_2)-\frac{1}{4\pi}\int_{-\infty}^{\infty}h(k',n,n,t)E_n\big(g_1,\tfrac12+it\big)\overline{E_n\big(g_2,\tfrac12+it\big)}\,dt$$

defines a compact operator $T_{K''}:S^n_{\mathrm{cusp}}\to S^n_{\mathrm{cusp}}$ that acts trivially on Θ. Note that

$$\int_{-\infty}^{\infty}h(k',n,n,t)E_n\big(g_1,\tfrac12+it\big)\overline{E_n\big(g_2,\tfrac12+it\big)}\,dt$$
$$=\int_{-\infty}^{\infty}\frac{h(k',n,n,t)}{\overline{c_{m,n}(1/2+it)}}E_n\big(g_1,\tfrac12+it\big)\overline{\boldsymbol{E}^{n-m}E_m\big(g_2,\tfrac12+it\big)}\,dt.$$

Let

$$K'''(g_1,g_2):=K_{m,n}(g_1,g_2)-\frac{1}{4\pi}\int_{-\infty}^{\infty}\frac{h(k',n,n,t)}{\overline{c_{m,n}(1/2+it)}}E_n\big(g_1,\tfrac12+it\big)\overline{E_m\big(g_2,\tfrac12+it\big)}\,dt.$$

Note that

$$T_{K''}=T_{K'''}\circ\boldsymbol{E}^{m-n}.$$

Firstly, since $\boldsymbol{E}^{m-n}$ does not annihilate the Eisenstein series, $T_{K'''}$ acts trivially on Θ.

If $m>n\geq 0$ or $m<n\leq 0$, then as a map $S^n_{\mathrm{cusp}}\to S^m_{\mathrm{cusp}}$, $ker(\boldsymbol{E}^{m-n})$ is empty, and we may decompose S^m_{cusp} as

$$S^m_{\mathrm{cusp}}=\Im(\boldsymbol{E}^{m-n})\oplus R,$$

where R is a finite dimensional subspace of S^m_{cusp} spanned by modular forms of weight $>n$ and their images under raising operators in S^m_{cusp}. Note that

$$(\boldsymbol{E}^{m-n})^{-1}:\Im(\boldsymbol{E}^{m-n})\to S^n_{\mathrm{cusp}}$$

is a bounded operator, hence

$$T_{K'''}|_{Im(\boldsymbol{E}^{m-n})} = T_{K''} \circ (\boldsymbol{E}^{m-n})^{-1}$$

is a compact operator. This implies that $T_{K'''}$ is a direct sum of a compact operator and finite dimensional linear operator, which is a compact operator.

If $n > m \geq 0$ or $n < m \leq 0$, then $\boldsymbol{E}^{m-n} : S^n_{\text{cusp}} \to S^m_{\text{cusp}}$ is surjective, and so we may define a bounded operator

$$(\boldsymbol{E}^{m-n})^{-1} : S_m \to (\ker(\boldsymbol{E}^{m-n}))^{\perp}$$

from which it follows that

$$T_{K'''} = T_{K''} \circ (\boldsymbol{E}^{m-n})^{-1}$$

is a compact operator.

If $n > 0 > m$ or $m > 0 > n$, then we further decompose $T_{K''}$ to

$$S^n_{\text{cusp}} \xrightarrow{\boldsymbol{E}^{-n}} S^0_{\text{cusp}} \xrightarrow{\boldsymbol{E}^{m}} S^m_{\text{cusp}} \xrightarrow{T_{K'''}} S^n_{\text{cusp}},$$

and then combine the above arguments to see that $T_{K''}$ is a compact operator.

Finally, observe that

$$h(k', n, n, t) = \int_{\mathrm{PSL}_2(\mathbb{R})} (\overline{\boldsymbol{E}^{n-m}} k(g)) y^{\frac{1}{2}+it} e^{in\theta}\, dg = c_{n,m}\left(\tfrac{1}{2}+it\right) \int_{\mathrm{PSL}_2(\mathbb{R})} k(g) y^{\frac{1}{2}+it} e^{im\theta}\, dg,$$

and we complete the proof using (5-6). □

5C. *General case.* We are now ready to describe Selberg's pretrace formula for $\mathrm{PSL}_2(\mathbb{R})$.

Theorem 5.8. *For $k \in C_0^{\infty}(\mathrm{PSL}_2(\mathbb{R}))$, let K be the integral kernel on $S\mathbb{X}$ defined by*

$$K(g_1, g_2) = \sum_{\gamma \in \Gamma} k(g_1, \gamma g_2).$$

Then we have

$$K(g_1, g_2) = \frac{9}{\pi^4} \iint K(g_1, g_2)\, dg_1\, dg_2 + \sum_{\substack{e \geq 0 \\ 2 \mid e}} \sum_{j=1}^{d_e} \sum_{\substack{m,n \in 2\mathbb{Z} \\ |m|,|n| \geq e}} h(k, m, n, \pi_j^e) \phi^e_{j,n}(g_1) \overline{\phi^e_{j,m}(g_2)}$$
$$+ \frac{1}{4\pi} \sum_{m,n \in 2\mathbb{Z}} \int_{-\infty}^{\infty} h(k, m, n, t) E_n\left(g_1, \tfrac{1}{2}+it\right) \overline{E_m\left(g_2, \tfrac{1}{2}+it\right)}\, dt,$$

where π_j^e is the irreducible unitary representation of $\mathrm{PSL}_2(\mathbb{R})$ associated to ϕ_j^e.

Proof. We first note from (5-2) and (5-5) that

$$\begin{aligned} K_{m,n}(g_1,g_2) &= \sum_{\gamma\in\Gamma} \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} k(R_{\theta_1'}g_1^{-1}\gamma g_2 R_{\theta_2'})e^{-in\theta_1'-im\theta_2'}\,d\theta_1'\,d\theta_2' \\ &= \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}\sum_{\gamma\in\Gamma} k(R_{-\theta_1'}g_1^{-1}\gamma g_2 R_{\theta_2'})e^{in\theta_1'-im\theta_2'}\,d\theta_1'\,d\theta_2' \\ &= \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} K(g_1R_{\theta_1'},g_2R_{\theta_2'})e^{in\theta_1'-im\theta_2'}\,d\theta_1'\,d\theta_2' \\ &= \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} K((x_1,y_1,\theta_1'),(x_2,y_2,\theta_2'))e^{in\theta_1'-im\theta_2'}\,d\theta_1'\,d\theta_2' e^{-in\theta_1+im\theta_2}. \end{aligned}$$

Therefore, we have the Fourier expansion of K,

$$K(g_1,g_2)=\sum_{n,m\in 2\mathbb{Z}} K_{m,n}(g_1,g_2),$$

where the summation is uniform for g_1 and g_2 in compacta.

We infer from Theorem 5.7 that

$$K_{m,n}(g_1,g_2)-\frac{1}{4\pi}\int_{-\infty}^{\infty} h(k,m,n,t)E_n\big(g_1,\tfrac12+it\big)\overline{E_m\big(g_2,\tfrac12+it\big)}\,dt$$

defines a compact operator acting on L_{cusp} that acts trivially on Θ. Because it only acts nontrivially on weight m forms, we see that

$$\begin{aligned} K_{m,n}(g_1,g_2)-\frac{1}{4\pi}\int_{-\infty}^{\infty} h(k,m,n,t)E_n\big(g_1,\tfrac12+it\big)\overline{E_m\big(g_2,\tfrac12+it\big)}\,dt \\ = \frac{9}{\pi^4}\iint K_{m,n}(g_1,g_2)\,dg_1\,dg_2+\sum_{\substack{e\ge 0\\ 2\mid e}}^{\min\{|m|,|n|\}}\sum_{j=1}^{d_e} h(k,m,n,\pi_j^e)\phi_{j,n}^e(g_1)\overline{\phi_{j,m}^e(g_2)}, \end{aligned}$$

where we used (5-1), and the fact that

$$\int_{-\infty}^{\infty} h(k,m,n,t)E_n\big(g_1,\tfrac12+it\big)\overline{E_m\big(g_2,\tfrac12+it\big)}\,dt$$

acts trivially on L^2_{cusp}. Note that the integral on the right-hand side of the equation vanishes unless $m=n=0$, in which case it is identical to

$$\frac{9}{\pi^4}\iint K(g_1,g_2)\,dg_1\,dg_2. \qquad \square$$

5D. ***Proof of Theorem 1.4.*** We now present a proof of Theorem 1.4. By Theorem 5.8, we have

$$\frac{1}{l(\mathscr{C}_{d_1})l(\mathscr{C}_{d_2})}\int_{\mathscr{C}_{d_2}}\int_{\mathscr{C}_{d_1}} K(s_1,s_2)\,ds_1\,ds_2 = M+D+\frac{1}{4\pi}E,$$

where

$$M = \frac{9}{\pi^4}\iint K(g_1, g_2)\, dg_1\, dg_2,$$

$$D = \sum_{\substack{e\ge 0\\ 2\mid e}}\sum_{j=1}^{d_e}\sum_{\substack{m,n\in 2\mathbb{Z}\\ |m|,|n|\ge e}} h(k, m, n, \pi_j^e)\frac{\mu_{d_1}(\phi_{j,n}^e)}{l(\mathscr{C}_{d_1})}\frac{\overline{\mu_{d_2}(\phi_{j,m}^e)}}{l(\mathscr{C}_{d_2})}$$

$$= \sum_{\substack{e\ge 0\\ 4\mid e}}\sum_{j=1}^{d_e}\frac{\mu_{d_1}(\phi_j^e)}{l(\mathscr{C}_{d_1})}\frac{\overline{\mu_{d_2}(\phi_j^e)}}{l(\mathscr{C}_{d_2})}\sum_{\substack{m,n\in 4\mathbb{Z}\\ |m|,|n|\ge e}} h(k, m, n, \pi_j^e)\eta_j^e(\phi_{j,n}^e)\overline{\eta_j^e(\phi_{j,m}^e)},$$

and

$$E = \sum_{m,n\in 2\mathbb{Z}}\int_{-\infty}^{\infty} h(k, m, n, t)\frac{\mu_{d_1}(E_n(\cdot\,, 1/2+it))}{l(\mathscr{C}_{d_1})}\frac{\overline{\mu_{d_2}(E_m(\cdot\,, 1/2+it))}}{l(\mathscr{C}_{d_2})}\, dt$$

$$= \sum_{m,n\in 4\mathbb{Z}}\int_{-\infty}^{\infty} h(k, m, n, t)\frac{\mu_{d_1}(E_0(\cdot\,, 1/2+it))}{l(\mathscr{C}_{d_1})}\frac{\overline{\mu_{d_2}(E_0(\cdot\,, 1/2+it))}}{l(\mathscr{C}_{d_2})}\eta\big(n, \tfrac{1}{2}+it\big)\overline{\eta\big(m, \tfrac{1}{2}+it\big)}dt.$$

For D with $e>0$, we use (4-2), (4-5), Lemma 5.3 with $N_1 = N_2 = 5$, and (4-10) to see that

$$\sum_{\substack{e>0\\ 4\mid e}}\sum_{j=1}^{d_e}\frac{\mu_{d_1}(\phi_j^e)}{l(\mathscr{C}_{d_1})}\frac{\overline{\mu_{d_2}(\phi_j^e)}}{l(\mathscr{C}_{d_2})}\sum_{\substack{m,n\in 4\mathbb{Z}\\ |m|,|n|\ge e}} h(k, m, n, \pi_j^e)\eta_j^e(\phi_{j,n}^e)\overline{\eta_j^e(\phi_{j,m}^e)}$$

$$\ll_\epsilon \sum_{\substack{e>0\\ 4\mid e}} e^{6.8}(d_1d_2)^{-25/512+\epsilon}\sum_{\substack{m,n\in 4\mathbb{Z}\\ |m|,|n|\ge e}} |m|^{-5}|n|^{-5}\|k\|_{W^{10,\infty}}$$

$$\ll (d_1d_2)^{-25/512+\epsilon}\|k\|_{W^{10,\infty}}.$$

For D with $e=0$, we use (4-1), (4-6), Lemma 5.5 with $N_1 = N_2 = 2$ and $N_3 = 4$, and (4-10) to see that

$$\sum_{j=1}^{\infty}\frac{\mu_{d_1}(\phi_j^0)}{l(\mathscr{C}_{d_1})}\frac{\overline{\mu_{d_2}(\phi_j^0)}}{l(\mathscr{C}_{d_2})}\sum_{m,n\in 4\mathbb{Z}} h(k, m, n, \pi_j^0)\eta_j^0(\phi_{j,n}^0)\overline{\eta_j^0(\phi_{j,m}^0)}$$

$$\ll_\epsilon \sum_{j=1}^{\infty}(d_1d_2)^{-25/512+\epsilon}|t_j|^{3/2}\sum_{m,n\in 4\mathbb{Z}}(1+|m|)^{-2}(1+|n|)^{-2}(1+|t_j|)^{-4}\|k\|_{W^{8,\infty}}$$

$$\ll (d_1d_2)^{-25/512+\epsilon}\|k\|_{W^{8,\infty}}.$$

For E, we use (4-3), (4-9), Lemma 5.5 with $N_1 = N_2 = 2$ and $N_3 = 3$, and (4-10) to see that

$$\sum_{m,n\in 4\mathbb{Z}}\int_{-\infty}^{\infty} h(k, m, n, t)\frac{\mu_{d_1}(E_0(\cdot\,, 1/2+it))}{l(\mathscr{C}_{d_1})}\frac{\overline{\mu_{d_2}(E_0(\cdot\,, 1/2+it))}}{l(\mathscr{C}_{d_2})}\eta\big(n, \tfrac{1}{2}+it\big)\overline{\eta\big(m, \tfrac{1}{2}+it\big)}\, dt$$

$$\ll_\epsilon \sum_{m,n\in 4\mathbb{Z}}\int_{-\infty}^{\infty}(d_1d_2)^{-1/16+\epsilon}(1+|m|)^{-2}(1+|n|)^{-2}(|t|+1)^{-2}\|k\|_{W^{7,\infty}}\, dt$$

$$\ll (d_1d_2)^{-1/16+\epsilon}\|k\|_{W^{7,\infty}}.$$

Now observe that

$$\iint K(g_1, g_2)\, dg_1\, dg_2 = \int_{S\mathbb{X}} \int_{S\mathbb{H}} k(g_1^{-1} g_2)\, dg_2\, dg_1 = \frac{\pi^2}{3} \int_{S\mathbb{H}} k(g)\, dg,$$

and so

$$M = \frac{3}{\pi^2} \int_{S\mathbb{H}} k(g)\, dg.$$

So far, we proved the following:

Theorem 5.9. *For any $k \in C_0^\infty(S\mathbb{H})$, we have*

$$\frac{1}{l(\mathscr{C}_{d_1}) l(\mathscr{C}_{d_2})} \int_{\mathscr{C}_{d_2}} \int_{\mathscr{C}_{d_1}} K(s_1, s_2)\, ds_1\, ds_2 = \frac{3}{\pi^2} \int_{S\mathbb{H}} k(g)\, dg + O_\epsilon((d_1 d_2)^{-25/512+\epsilon} \|k\|_{W^{10,\infty}}).$$

Remark 5.1. Note that this is *not* the same as equidistribution of $\mathscr{C}_{d_1} \times \mathscr{C}_{d_2}$ in $S\mathbb{X} \times S\mathbb{X}$. For instance, if we replace K with any compactly supported smooth function in $S\mathbb{X} \times S\mathbb{X}$, then the equality may not hold when d_1 is fixed and d_2 tends to ∞.

In order to prove Theorem 1.4, we make specific choices of k in Theorem 5.9. We let K_1 and K_2 to be the kernel corresponding to $k = m_\delta^{\theta_1,\theta_2}$ and $k = M_\delta^{\theta_1,\theta_2}$ defined in Lemma 2.4, respectively. Then by Lemma 2.4, we have

$$\begin{aligned} \frac{1}{l(\mathscr{C}_{d_1}) l(\mathscr{C}_{d_2})} \int_{\mathscr{C}_{d_2}} \int_{\mathscr{C}_{d_1}} K_1(s_1, s_2)\, ds_1\, ds_2 &\le \frac{1}{l(\mathscr{C}_{d_1}) l(\mathscr{C}_{d_2})} \int_{\mathscr{C}_{d_2}} \int_{\mathscr{C}_{d_1}} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 \\ &\le \frac{1}{l(\mathscr{C}_{d_1}) l(\mathscr{C}_{d_2})} \int_{\mathscr{C}_{d_2}} \int_{\mathscr{C}_{d_1}} K_2(s_1, s_2)\, ds_1\, ds_2, \end{aligned}$$

while we know from Lemma 2.6 that

$$\int_{\mathscr{C}_{d_2}} \int_{\mathscr{C}_{d_1}} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 = 4\delta^2 I_{\theta_1,\theta_2}(C_{d_1}, C_{d_2}).$$

We now apply Theorem 5.9 and Lemma 2.4 to see that

$$\frac{1}{l(\mathscr{C}_{d_1}) l(\mathscr{C}_{d_2})} \int_{\mathscr{C}_{d_2}} \int_{\mathscr{C}_{d_1}} K_i(s_1, s_2)\, ds_1\, ds_2 = \frac{3}{\pi^2} (\cos\theta_1 - \cos\theta_2)\delta^2 (1 + O(\varepsilon)) + O_\epsilon((d_1 d_2)^{-25/512+\epsilon} \varepsilon^{-10}).$$

Therefore, we have

$$\frac{I_{\theta_1,\theta_2}(C_{d_1}, C_{d_2})}{l(C_{d_1}) l(C_{d_2})} = \frac{3}{\pi^2} (\cos\theta_1 - \cos\theta_2)(1 + O(\delta^2))(1 + O(\varepsilon)) + O_\epsilon((d_1 d_2)^{-25/512+\epsilon} \varepsilon^{-10} \delta^{-2}),$$

and by choosing $\delta^2 = \varepsilon = (d_1 d_2)^{-25/6144}$, we complete the proof of Theorem 1.4.

Appendix: Jacobian computation

Recall that $\Psi : AKA \to \mathrm{SL}_2(\mathbb{R})$ is given by

$$(t_1, \varphi, t_2) \mapsto \begin{pmatrix} e^{t_1/2} & 0 \\ 0 & e^{-t_1/2} \end{pmatrix} R_{\frac{\varphi}{2}} \begin{pmatrix} e^{-t_2/2} & 0 \\ 0 & e^{t_2/2} \end{pmatrix} = \begin{pmatrix} e^{(t_1-t_2)/2} \cos\frac{\varphi}{2} & -e^{(t_1+t_2)/2} \sin\frac{\varphi}{2} \\ e^{(-t_1-t_2)/2} \sin\frac{\varphi}{2} & e^{(t_2-t_1)/2} \cos\frac{\varphi}{2} \end{pmatrix}.$$

In this section, we compute the pullback of $dV = dx\,dy\,d\theta/y^2$ under Ψ. We start with the identity

$$\begin{pmatrix} e^{(t_1-t_2)/2}\cos\frac{\varphi}{2} & -e^{(t_1+t_2)/2}\sin\frac{\varphi}{2} \\ e^{(-t_1-t_2)/2}\sin\frac{\varphi}{2} & e^{(t_2-t_1)/2}\cos\frac{\varphi}{2} \end{pmatrix} = n(x)a(y)R_\theta = \begin{pmatrix} * & * \\ \frac{\sin\theta}{\sqrt{y}} & \frac{\cos\theta}{\sqrt{y}} \end{pmatrix}.$$

By comparing the image of $i \in \mathbb{H}$, we have

$$x+iy = \frac{e^{(t_1-t_2)/2}\cos\frac{\varphi}{2}i - e^{(t_1+t_2)/2}\sin\frac{\varphi}{2}}{e^{(-t_1-t_2)/2}\sin\frac{\varphi}{2}i + e^{(t_2-t_1)/2}\cos\frac{\varphi}{2}},$$

and for simplicity, we write this as $\frac{A}{B}$. By comparing the second row of each matrix, we have

$$\frac{e^{i\theta}}{\sqrt{y}} = B.$$

From a quick computation, we see that

$$A_{t_1} = \frac{A}{2},\quad B_{t_1} = -\frac{B}{2},\quad A_{t_2} = \frac{\bar{A}}{2},\quad B_{t_2} = \frac{\bar{B}}{2},\quad A_\varphi = -\frac{e^{t_1}}{2}B,\quad B_\varphi = \frac{e^{-t_1}}{2}A,\quad \Im A\bar{B} = 1,\quad y = \frac{1}{|B|^2}.$$

We use these to express the Jacobian matrix in terms of A and B as follows:

$$\frac{\partial(x,y,\theta)}{\partial(t_1,t_2,\varphi)} = \begin{pmatrix} \operatorname{Re}\frac{A}{B} & \Im\frac{1}{B^2} & \operatorname{Re}\big(-\frac{e^{t_1}}{2} - \frac{e^{-t_1}}{2}\frac{A^2}{B^2}\big) \\ \Im\frac{A}{B} & -\operatorname{Re}\frac{1}{B^2} & \Im\big(-\frac{e^{t_1}}{2} - \frac{e^{-t_1}}{2}\frac{A^2}{B^2}\big) \\ 0 & \frac{1}{2}\Im\frac{\bar{B}}{B} & \frac{e^{-t_1}}{2|B|^2} \end{pmatrix}.$$

From this, we have

$$\begin{aligned} \frac{1}{y^2}\left|\frac{\partial(x,y,\theta)}{\partial(t_1,t_2,\varphi)}\right| &= |B|^4\left|\frac{\partial(x,y,\theta)}{\partial(t_1,t_2,\varphi)}\right| \\ &= \left|-\tfrac{1}{2}e^{-t_1}\operatorname{Re}\left(\frac{\bar{A}}{B}\right) + \tfrac{1}{4}\Im(\bar{B}^2)\Im\left(\bar{A}B\left(e^{t_1} + e^{-t_1}\frac{A^2}{B^2}\right)\right)\right| \\ &= \left|\frac{e^{t_1}}{2}\Im(B^2) + \frac{e^{-t_1}}{4|B|^2}(-2\operatorname{Re}(AB) - |A|^2\Im(B^2))\right|. \end{aligned}$$

Now we use the definition of A and B to compute each term explicitly as follows

$$\begin{aligned} 2\operatorname{Re}(AB) &= -(e^{t_2} + e^{-t_1})\sin\varphi \\ e^{t_1}\Im(B^2) &= \sin\varphi \\ e^{-t_1}|A|^2 &= e^{t_2}\sin^2\frac{\varphi}{2} + e^{-t_2}\cos^2\frac{\varphi}{2} \\ e^{t_1}|B|^2 &= e^{-t_2}\sin^2\frac{\varphi}{2} + e^{t_2}\cos^2\frac{\varphi}{2}, \end{aligned}$$

and so

$$\frac{1}{y^2}\left|\frac{\partial(x,y,\theta)}{\partial(t_1,t_2,\varphi)}\right| = \tfrac{1}{2}|\sin\varphi|.$$

Therefore, we conclude that

$$dV = \tfrac{1}{2}|\sin\varphi|\,dt_1\,dt_2\,d\varphi. \tag{A-1}$$

References

[Artin 1924] E. Artin, "Ein mechanisches System mit quasiergodischen Bahnen", *Abh. Math. Sem. Univ. Hamburg* **3**:1 (1924), 170–175. MR Zbl

[Baruch and Mao 2010] E. M. Baruch and Z. Mao, "A generalized Kohnen–Zagier formula for Maass forms", *J. Lond. Math. Soc.* (2) **82**:1 (2010), 1–16. MR Zbl

[Blomer et al. 2007] V. Blomer, G. Harcos, and P. Michel, "A Burgess-like subconvex bound for twisted L-functions", *Forum Math.* **19**:1 (2007), 61–105. MR Zbl

[Bonahon 1986] F. Bonahon, "Bouts des variétés hyperboliques de dimension 3", *Ann. of Math.* (2) **124**:1 (1986), 71–158. MR Zbl

[Darmon and Vonk 2022] H. Darmon and J. Vonk, "Arithmetic intersections of modular geodesics", *J. Number Theory* **230** (2022), 89–111. MR Zbl

[Davenport 1967] H. Davenport, *Multiplicative number theory*, Lectures in Advanced Mathematics **1**, Markham Publishing Co., Chicago, 1967. MR Zbl

[Duke 1988] W. Duke, "Hyperbolic distribution problems and half-integral weight Maass forms", *Invent. Math.* **92**:1 (1988), 73–90. MR Zbl

[Heath-Brown 1980] D. R. Heath-Brown, "Hybrid bounds for Dirichlet L-functions, II", *Quart. J. Math. Oxford Ser.* (2) **31**:122 (1980), 157–167. MR Zbl

[Hejhal 1976] D. A. Hejhal, *The Selberg trace formula for* $\mathrm{PSL}(2, R)$, *I*, Lecture Notes in Mathematics **548**, Springer, 1976. MR Zbl

[Herrera Jaramillo 2015] Y. A. Herrera Jaramillo, "Intersection numbers of geodesic arcs", *Rev. Colombiana Mat.* **49**:2 (2015), 307–319. MR

[Jakobson 1994] D. Jakobson, "Quantum unique ergodicity for Eisenstein series on $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}_2(\mathbb{R})$", *Ann. Inst. Fourier (Grenoble)* **44**:5 (1994), 1477–1504. MR Zbl

[Jung and Reid 2021] J. Jung and A. W. Reid, "Embedding closed totally geodesic surfaces in Bianchi orbifolds", preprint, 2021. To appear in *Math. Res. Lett.* arXiv 2003.05427

[Katok 1985] S. Katok, "Closed geodesics, periods and arithmetic of modular forms", *Invent. Math.* **80**:3 (1985), 469–480. MR Zbl

[Katok and Sarnak 1993] S. Katok and P. Sarnak, "Heegner points, cycles and Maass forms", *Israel J. Math.* **84**:1-2 (1993), 193–227. MR Zbl

[Knapp 2001] A. W. Knapp, *Representation theory of semisimple groups*, Princeton University Press, 2001. MR Zbl

[Kohnen and Zagier 1981] W. Kohnen and D. Zagier, "Values of L-series of modular forms at the center of the critical strip", *Invent. Math.* **64**:2 (1981), 175–198. MR Zbl

[Kubota 1973] T. Kubota, *Elementary theory of Eisenstein series*, Kodansha, Ltd., Tokyo, 1973. MR Zbl

[Lalley 2014] S. P. Lalley, "Statistical regularities of self-intersection counts for geodesics on negatively curved surfaces", *Duke Math. J.* **163**:6 (2014), 1191–1261. MR Zbl

[Lang 1985] S. Lang, $\mathrm{SL}_2(\mathbf{R})$, Graduate Texts in Mathematics **105**, Springer, 1985. MR

[Lester and Radziwiłł 2020] S. Lester and M. Radziwiłł, "Quantum unique ergodicity for half-integral weight automorphic forms", *Duke Math. J.* **169**:2 (2020), 279–351. MR Zbl

[Luo et al. 2009] W. Luo, Z. Rudnick, and P. Sarnak, "The variance of arithmetic measures associated to closed geodesics on the modular surface", *J. Mod. Dyn.* **3**:2 (2009), 271–309. MR Zbl

[Petrow and Young 2019] I. Petrow and M. P. Young, "A generalized cubic moment and the Petersson formula for newforms", *Math. Ann.* **373**:1-2 (2019), 287–353. MR Zbl

[Pollicott and Sharp 2006] M. Pollicott and R. Sharp, "Angular self-intersections for closed geodesics on surfaces", *Proc. Amer. Math. Soc.* **134**:2 (2006), 419–426. MR Zbl

[Rickards 2021] J. Rickards, "Computing intersections of closed geodesics on the modular curve", *J. Number Theory* **225** (2021), 374–408. MR Zbl

[Rivin 2001] I. Rivin, "Simple curves on surfaces", *Geom. Dedicata* **87**:1-3 (2001), 345–360. MR Zbl

[Sardari 2021] N. T. Sardari, "The least prime number represented by a binary quadratic form", *J. Eur. Math. Soc. (JEMS)* **23**:4 (2021), 1161–1223. MR Zbl

[Sarnak 1980] P. C. Sarnak, *Prime geodesic theorems*, Ph.D. thesis, Stanford University, 1980, https://www.proquest.com/docview/303065936. MR

[Sarnak 1982] P. Sarnak, "Class numbers of indefinite binary quadratic forms", *J. Number Theory* **15**:2 (1982), 229–247. MR Zbl

[Shintani 1975] T. Shintani, "On construction of holomorphic cusp forms of half integral weight", *Nagoya Math. J.* **58** (1975), 83–126. MR Zbl

[Siegel 1935] C. L. Siegel, "Über die Classenzahl quadratischer Zahlkörper", *Acta Arith.* **1**:1 (1935), 83–36. Zbl

[Siegel 1965] C. L. Siegel, *Lectures on advanced analytic number theory*, Tata Institute of Fundamental Research Lectures on Mathematics **23**, Tata Institute of Fundamental Research, Bombay, 1965. MR Zbl

[Zagier 1981] D. Zagier, "Eisenstein series and the Riemann zeta function", pp. 275–301 in *Automorphic forms, representation theory and arithmetic* (Bombay, 1979), edited by K. G. Ramanathan, Tata Inst. Fund. Res. Studies in Math. **10**, Tata Institute of Fundamental Research, Bombay, 1981. MR Zbl

Communicated by Philippe Michel
Received 2021-12-09 Revised 2022-05-18 Accepted 2022-07-06

junehyuk_jung@brown.edu Department of Mathematics, Brown University, Providence, RI, United States

nzt5208@psu.edu Department of Mathematics, Pennsylvania State University, State College, PA, United States

Algebra & Number Theory

msp.org/ant

EDITORS

Managing Editor	Editorial Board Chair
Antoine Chambert-Loir	David Eisenbud
Université Paris-Diderot	University of California
France	Berkeley, USA

Board of Editors

Jason P. Bell	University of Waterloo, Canada	Philippe Michel	École Polytechnique Fédérale de Lausanne
Bhargav Bhatt	University of Michigan, USA	Martin Olsson	University of California, Berkeley, USA
Frank Calegari	University of Chicago, USA	Irena Peeva	Cornell University, USA
J-L. Colliot-Thélène	CNRS, Université Paris-Saclay, France	Jonathan Pila	University of Oxford, UK
Brian D. Conrad	Stanford University, USA	Anand Pillay	University of Notre Dame, USA
Samit Dasgupta	Duke University, USA	Bjorn Poonen	Massachusetts Institute of Technology, USA
Hélène Esnault	Freie Universität Berlin, Germany	Victor Reiner	University of Minnesota, USA
Gavril Farkas	Humboldt Universität zu Berlin, Germany	Peter Sarnak	Princeton University, USA
Sergey Fomin	University of Michigan, USA	Michael Singer	North Carolina State University, USA
Edward Frenkel	University of California, Berkeley, USA	Vasudevan Srinivas	Tata Inst. of Fund. Research, India
Wee Teck Gan	National University of Singapore	Shunsuke Takagi	University of Tokyo, Japan
Andrew Granville	Université de Montréal, Canada	Pham Huu Tiep	Rutgers University, USA
Ben J. Green	University of Oxford, UK	Ravi Vakil	Stanford University, USA
Christopher Hacon	University of Utah, USA	Akshay Venkatesh	Institute for Advanced Study, USA
Roger Heath-Brown	Oxford University, UK	Melanie Matchett Wood	Harvard University, USA
János Kollár	Princeton University, USA	Shou-Wu Zhang	Princeton University, USA
Michael J. Larsen	Indiana University Bloomington, USA		

PRODUCTION

production@msp.org

Silvio Levy, Scientific Editor

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2023 is US $485/year for the electronic version, and $705/year (+$65, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online.

ANT peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

© 2023 Mathematical Sciences Publishers

Algebra & Number Theory

Volume 17 No. 7 2023