

Equidistribution of q -orbits of closed geodesics

Asbjørn Christian Nordentoft

Abstract. Given an element of $(\mathbb{Z}/q\mathbb{Z})^\times$ we associate a closed geodesic on the modular surface. We prove that the closed geodesics associated to cosets of sufficiently large subgroups equidistribute in the unit tangent bundle as q tends to infinity. This is a q -orbit analogue of Duke’s theorem for real quadratic fields as extended to subgroups by Popa. Finally, we show that the homology classes of the q -orbits of oriented closed geodesics concentrate around the Eisenstein line and present group theoretic applications.

1. Introduction

The modular surface $Y_0(1) := \mathrm{PSL}_2(\mathbb{Z}) \backslash \mathbb{H}$ is naturally endowed with the structure of a non-compact Riemannian orbifold with line and volume elements given by

$$ds = \frac{|dz|}{\mathrm{Im} z} \quad \text{and} \quad d\mu = \frac{-dz d\bar{z}}{2i(\mathrm{Im} z)^2},$$

in terms of the standard coordinate $z \in \mathbb{H}$. The unit tangent bundle admits a natural description as a homogeneous space

$$\mathbf{T}^1(Y_0(1)) = \mathrm{PSL}_2(\mathbb{Z}) \backslash \mathrm{PSL}_2(\mathbb{R}),$$

which we equip with the Haar probability measure dg . The distribution of closed geodesics on the modular surface is a classical and rich topic starting with work of Linnik [35] and Skubenko [48] who proved equidistribution in the unit tangent bundle for individual length packets of closed geodesics under certain congruence conditions. These congruence conditions were later removed in the celebrated work of Duke [15]. More precisely, let $D > 0$ be a positive discriminant with narrow class group Cl_D^+ . Given a narrow ideal class $B \in \mathrm{Cl}_D^+$ there is an associated oriented closed geodesic $\mathcal{C}_B \subset \mathbf{T}^1(Y_0(1))$ in the unit tangent bundle of the modular surface with geodesic length

$$|\mathcal{C}_B| = 2 \log \epsilon_D,$$

Mathematics Subject Classification 2020: 11F67.

Keywords: equidistribution, closed geodesics, modular surface, L -functions.

where ϵ_D denotes the fundamental positive unit of discriminant D ; see [44]. The closed geodesic $\mathcal{C}_B$ is a closed A -orbit where A denotes the diagonal subgroup of $\mathrm{PSL}_2(\mathbb{R})$. We equip $\mathcal{C}_B$ with the unique A -invariant measure μ_B , which descends to ds on $Y_0(1)$. Duke's theorem states that for any smooth and compactly supported $\phi: \mathbf{T}^1(Y_0(1)) \rightarrow \mathbb{C}$, we have

$$\frac{\sum_{B \in \mathrm{Cl}_D^+} \int_{\mathcal{C}_B} \phi(t) d\mu_B(t)}{2|\mathrm{Cl}_D^+| \log \epsilon_D} \rightarrow \int_{\mathbf{T}^1(Y_0(1))} \phi(g) dg \quad \text{as } D \rightarrow \infty.$$

It is natural to ask to what extent other collections of closed geodesics equidistribute. This is a special case of the *Basic Question* posed in [18, Section 1.1]. In this context, Michel and Venkatesh, in their ICM address [38], put forth a number of “sparse” equidistribution conjectures refining Duke's theorem. Progress towards these conjectures include the works of Harcos–Michel [23] and Popa [43, Theorem 6.5.1] obtaining equidistribution for closed geodesics associated to cosets of sufficiently large subgroups of class groups Cl_D^+ (of index $\leq D^\delta$ for some $\delta > 0$) and the work of Einsiedler–Lindenstrauss–Michel–Venkatesh [19] obtaining equidistribution of sufficiently large subcollections (of index $\leq D^{o(1)}$) using ergodic techniques; see also [1]. The Mixing conjecture [38, Conjecture 2] has also seen recent progress; see [2, 5, 32]. In a different direction, Erlandsson–Souto [20] proved equidistribution for the special set of reciprocal geodesics answering a question of Sarnak [45]. In the other extreme, Bourgain–Kontorovich [7] showed the existence of many ($\gg_\epsilon D^{1/2-\epsilon}$ for any $\epsilon > 0$) low-lying closed fundamental geodesics which in particular do not equidistribute.

In this paper, we introduce “ q -orbits” of closed geodesics in analogy with the imaginary quadratic case of Heegner points (see Remark 1.2 below). We obtain an analogue of Duke's theorem (and its extension to subgroups due to Harcos–Michel [23] and Popa [43]), as well as concentration results for the homology classes of q -orbits of closed geodesics as studied in the setting of Duke's theorem by the author [40]. In view of the *Basic Question* [18, Section 1.1], our results can be interpreted as saying that the q -orbit structure on the closed geodesics does not conspire with the low-lying geodesics. The key technical result proved in this paper, Proposition 4.2, is a quantitative relation between geodesic periods and vertical periods of automorphic forms, which might be of independent interest.

1.1. Equidistribution of q -orbits of closed geodesics

To state our results, denote by $\Gamma_0(N) \subset \mathrm{PSL}_2(\mathbb{Z})$ the Hecke congruence subgroup of level $N \geq 1$ and by $Y_0(N) := \Gamma_0(N) \backslash \mathbb{H}$ the modular surface of level N equipped with the line element ds and volume element $d\mu$ as above. Let $q \geq 2$ be an integer

and consider an embedding (of sets)

$$\psi: (\mathbb{Z}/q\mathbb{Z})^\times \hookrightarrow \Gamma_0(q) \subset \mathrm{PSL}_2(\mathbb{Z}),$$

satisfying

- (1) $\psi(a)\infty \equiv (a/q) \bmod 1$ (i.e. $\psi(a)$ is of the shape $\begin{pmatrix} a & * \\ q & * \end{pmatrix}$),
- (2) $|\mathrm{tr}(\psi(a))| > 2$ (i.e. $\psi(a)$ is a hyperbolic matrix).

Notice that for $N|q$ we have $\psi(a) \in \Gamma_0(N)$, and we denote by

$$\mathcal{C}_a^\psi(N) \subset \mathbf{T}^1(Y_0(N)) = \Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R}),$$

the oriented closed geodesic associated to the conjugacy class of $\psi(a)$ inside $\Gamma_0(N)$ (see Section 2.1 for details on this correspondence). This way, we obtain the following q -orbit of oriented closed geodesics:

$$\{\mathcal{C}_a^\psi(N) \subset \mathbf{T}^1(Y_0(N)) : a \in (\mathbb{Z}/q\mathbb{Z})^\times\}, \quad (1.1)$$

by which we mean a “natural family” of closed geodesics parametrized by the multiplicative group $(\mathbb{Z}/q\mathbb{Z})^\times$ (in our case the parametrization stems from the double coset decomposition; see Section 2.1.1). We equip $\mathcal{C}_a^\psi(N)$ with the unique A -invariant measure μ_a^ψ , which descends to the line element ds when $\mathcal{C}_a^\psi(N)$ is projected to $Y_0(N)$; see the introduction of [53] for details. Notice that in any q -orbit each geodesic length is attained at most twice.

Our first result is the following equidistribution result in the unit tangent bundle for closed geodesics associated to cosets of sufficiently large subgroups using the parametrization of the q -orbits in terms of $(\mathbb{Z}/q\mathbb{Z})^\times$.

Theorem 1.1. *Fix $N \geq 1$, $\delta > 7/8$ and $M \geq 1$. For every $q \geq 2$ such that $N|q$, let*

$$\psi: (\mathbb{Z}/q\mathbb{Z})^\times \hookrightarrow \Gamma_0(N),$$

be an embedding satisfying the conditions (1) and (2), as well as $|\mathrm{tr}(\psi(a))| \leq Mq$ for all $a \in (\mathbb{Z}/q\mathbb{Z})^\times$, and let $cH \subset (\mathbb{Z}/q\mathbb{Z})^\times$ be a coset of a subgroup $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$ of size $\geq q^\delta$. Then for any smooth and compactly supported function $\phi: \mathbf{T}^1(Y_0(N)) \rightarrow \mathbb{C}$ and $\varepsilon > 0$, we have that

$$\begin{aligned} & \frac{\sum_{a \in cH} \int_{\mathcal{C}_a^\psi(N)} \phi(t) d\mu_a^\psi(t)}{\sum_{a \in cH} |\mathcal{C}_a^\psi(N)|} \\ &= \int_{\mathbf{T}^1(Y_0(N))} \phi(g) dg + O_{\phi, \varepsilon}((\log q)^{-1+\varepsilon}) \quad \text{as } q \rightarrow \infty. \end{aligned} \quad (1.2)$$

Here, $|\mathcal{C}|$ for a curve $\mathcal{C} \subset Y_0(N)$, denotes the arc length with respect to the hyperbolic line element ds and dg denotes the Haar probability measure on $\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R})$.

Remark 1.2. The study of q -orbits (or p -orbits if $q = p$ is prime) is a very natural one with the most elementary example being the equidistribution of residues and their inverses on the torus:

$$\left\{ \left(\frac{a}{q}, \frac{\bar{a}}{q} \right) : a \in (\mathbb{Z}/q\mathbb{Z})^\times \right\} \subset (\mathbb{R}/\mathbb{Z})^2.$$

We refer to [26] for sparse equidistribution results in this setting. Another classical example is the equidistribution of Hecke orbits (see e.g. [21, Section 2.3]). To explain this, we observe that for a prime p we have a natural parametrization of the Hecke operator T_p in terms of $\mathbf{P}^1(\mathbb{F}_p)$ given by

$$(T_p f)(g) = \frac{1}{p^{1/2}} \sum_{t \in \mathbf{P}^1(\mathbb{F}_p)} f(\gamma_t g), \quad f: \mathrm{PSL}_2(\mathbb{Z}) \backslash \mathrm{PSL}_2(\mathbb{R}) \rightarrow \mathbb{C},$$

where

$$\gamma_\infty = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}, \quad \gamma_t = \begin{pmatrix} 1 & t \\ 0 & p \end{pmatrix}, \quad t \in \mathbb{F}_p.$$

The collection of points

$$\{\gamma_t i : t \in \mathbf{P}^1(\mathbb{F}_p)\} \subset \mathrm{PSL}_2(\mathbb{Z}) \backslash \mathbb{H}, \quad (1.3)$$

we define as the p -orbit of the point $i \in \mathbb{H}$. Note that all but one of the points (1.3) lie on the horocycle

$$\left\{ x + \frac{i}{p} : 0 \leq x \leq 1 \right\}.$$

Any non-trivial bound for the Hecke eigenvalues implies the well-known equidistribution of p -orbits; the probability measures with point masses at the points (1.3) converge weakly to the normalized hyperbolic measure $\frac{3}{\pi} \frac{dx dy}{y^2}$ as $p \rightarrow \infty$. This is, in fact, closely related to the situation studied in the present paper: Theorem 1.1 in the case where $H = (\mathbb{Z}/q\mathbb{Z})^\times$ is the full multiplicative group reduces (morally at least) to a non-trivial bound of Hecke eigenvalues. More precisely, the Weyl sums in our setting are (essentially) obtained by integrating the Maaß forms along $\{\gamma_t i \mathbb{R} : t \in \mathbf{P}^1(\mathbb{F}_p)\}$ (see Remark 7.3).

Recall [47, Section 4.2] that the modular surface $\mathrm{PSL}_2(\mathbb{Z}) \backslash \mathbb{H}$ can be viewed as the moduli space for elliptic curves over $\mathbb{C}$ and that elliptic curves with complex multiplication (CM) correspond exactly to the *Heegner points* as defined in e.g. [11, Section 2.2]. From this perspective, we note as in [5, Remark 1.5] that the collection of points (1.3) coincides, after removing $\gamma_t i$ for $t^2 \equiv -1 \pmod{p}$, with the set of Heegner points associated to the non-fundamental discriminant $-4p^2$. One can now study all sorts of sparse equidistribution questions using the group structure on $\mathbb{F}_p^\times \subset \mathbf{P}^1(\mathbb{F}_p)$ (i.e. using the p -orbit structure). Notice however that this group

structure does not seem to admit a natural description in terms of moduli of CM elliptic curves since $\gamma_t i$ and $\gamma_{-t-1} i$ parametrize the same elliptic curve, but in general $\gamma_{at} i$ and $\gamma_{-at-1} i$ with $a \in \mathbb{F}_p^\times$ do not. Fouvry–Kowalski–Michel [21, Section 2.3] studied the equidistribution problem of p -orbits with certain algebraic weights (trace functions) associated to each of the Hecke points using the parametrization in terms of $\mathbf{P}^1(\mathbb{F}_p)$. An example being the indicator function of some coset of a subgroup of $\mathbb{F}_p^\times$ of sufficiently small index which is a direct analogue of Theorem 1.1. Recently, Blomer and Michel [5] resolved the analogue of the Mixing conjecture of Michel–Venkatesh [38] for p -orbits (unipotent mixing). It would be interesting to see if their methods extends to the setting of this paper.

1.2. The homology classes associated to closed geodesics

As investigated in [40] there is another natural distribution problem that one can study associated to oriented closed geodesics, namely the distribution of their homology classes. To set this up, let

$$V_N := H_1(Y_0(N), \mathbb{R}),$$

denote the singular homology group with $\mathbb{R}$ -coefficients of the (open) modular surface of level N . Given an oriented closed curve $\mathcal{C} \subset Y_0(N)$, we denote by

$$[\mathcal{C}] \in H_1(Y_0(N), \mathbb{Z}) \subset V_N$$

the associated homology class. Given $q \geq 2$ such that $N|q$ and an embedding

$$\psi: (\mathbb{Z}/q\mathbb{Z})^\times \hookrightarrow \Gamma_0(q),$$

satisfying the conditions (1) and (2), we get a map

$$(\mathbb{Z}/q\mathbb{Z})^\times \rightarrow V_N, \quad a \mapsto [\mathcal{C}_a^\psi(N)].$$

This map will depend very subtly on the sign of the traces of the image of ψ and thus we will furthermore assume that the absolute trace is *minimal* (see equation (2.12) and the discussion after Corollary 7.5 for further details):

(3) the absolute trace $|\operatorname{tr} \psi(a)|$ is minimal¹ subject to the conditions (1) and (2).

In order to understand the distribution inside the vector space V_N , we will have to normalize our classes. A convenient way to do this is to use the natural projection map

$$V_N - \{0\} \twoheadrightarrow \mathbf{S}(V_N) := (V_N - \{0\})/\mathbb{R}_{>0}, \quad v \mapsto \bar{v}.$$

¹For $a \in (\mathbb{Z}/q\mathbb{Z})^\times$ such that $a + a^{-1} \notin \{-2, -1, 0, 1, 2\} \bmod q$ this means exactly that $|\operatorname{tr} \psi(a)| \leq q/2$.

We equip $\mathbf{S}(V_N)$ with the quotient topology, which makes it a topological $2g$ -dimensional sphere. Finally, denote by

$$v_E(N) := [(x + i : 0 \leq x \leq 1)] \in H_1(Y_0(N), \mathbb{Z}),$$

the class of a simple loop going around the cusp at ∞ . It is a general fact that we have an isomorphism

$$H_1(Y_0(N), \mathbb{Z}) \xrightarrow{\sim} \Gamma_0(N)^{\text{ab}} / (\Gamma_0(N)^{\text{ab}})_{\text{tor}}, \quad (1.4)$$

which identifies $v_E(N)$ with the image of the matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ in the abelinization.

Our second result is the following concentration result for homology classes of closed geodesics associated to subgroups of $(\mathbb{Z}/q\mathbb{Z})^\times$.

Theorem 1.3. *Fix $N \geq 3$ prime. For each $q \geq 2$ such that $N|q$, choose an embedding*

$$\psi: (\mathbb{Z}/q\mathbb{Z})^\times \hookrightarrow \Gamma_0(q),$$

satisfying (1)–(3), and a subgroup $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$ such that $-1 \notin H$ and

$$\{a \in (\mathbb{Z}/q\mathbb{Z})^\times : a \equiv 1 \pmod{q/N}\} \not\subset H.$$

Then, we have

$$\overline{\sum_{a \in H} [\mathcal{C}_a^\psi(N)]} \rightarrow \overline{v_E(N)} \quad \text{as } q \rightarrow \infty$$

in the quotient topology of $\mathbf{S}(V_N)$.

Furthermore, we obtain a version of Theorem 1.3 which is uniform in the level N (see Theorem 7.4), which is a q -orbit analogue of the result of Liu–Masri–Young [37, Theorem 1.1] in the context of supersingular reduction of CM elliptic curves.

Finally, we will present the following group theoretic interpretation and application of Theorem 1.3. We assume for simplicity that N is prime and $N \equiv -1 \pmod{12}$, so that $\Gamma_0(N)$ is a free group on $2g + 1$ generators with $g = N/12 + O(1)$ (this follows from the discussion in [29, Section 2.3]). Then, as recently shown by Doan–Kim–Lang–Tan [13], one can find a set of free generators contained in

$$\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} a & -(aa^* + 1)/N \\ N & -a^* \end{pmatrix} : 0 < a < N \right\}, \quad (1.5)$$

where $0 < a^* < N$ is such that $aa^* \equiv -1 \pmod{N}$ (here we consider the matrices (1.5) as elements of $\text{PSL}_2(\mathbb{R})$ in the obvious way). It is natural to try to understand the representation of matrices in terms of such generating sets. If we look at a random word in the generators then we expect that there is about the same appearance of each generator. Theorem 1.3 implies that the q -orbits of closed geodesics are *not* random in

this sense since one coordinate dominates (see also Theorem 7.4 below). Geometrically this means that the q -orbits of closed geodesics “wind around the cusp a lot”. In particular, we obtain the following elementary group theoretic non-containment result, which seems hard to prove by elementary means (see [22] on Sarnak’s *Optimal lifting property* for another application of automorphic forms to group theory).

Corollary 1.4. *Let $N \equiv -1 \pmod{12}$ be prime and let*

$$\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \sigma_1, \dots, \sigma_{2g} \right\},$$

be a subset of (1.5), which define a set of free generators of $\Gamma_0(N)$. Then for each $\delta > 18$, there exists a constant $c(\delta) > 0$ such that the following holds: Let $q \geq c(\delta)N^\delta$ be an integer such that $N|q$ and let $0 < a_1, \dots, a_m < q - 1$ be pairwise distinct integers coprime to q such that

$$\{a_i \bmod q : 1 \leq i \leq m\} \subset (\mathbb{Z}/q\mathbb{Z})^\times$$

is a subgroup not containing $\{a \in (\mathbb{Z}/q\mathbb{Z})^\times : a \equiv 1 \pmod{q/N}\}$ (nor -1). For $1 \leq i \leq m$, let $0 < a_i^ < q$ be defined by $q|a_i a_i^* + 1$. Then no conjugate of*

$$\begin{pmatrix} a_1 & -\frac{a_1 a_1^* + 1}{q} \\ q & -a_1^* \end{pmatrix} \begin{pmatrix} a_2 & -\frac{a_2 a_2^* + 1}{q} \\ q & -a_2^* \end{pmatrix} \dots \begin{pmatrix} a_m & -\frac{a_m a_m^* + 1}{q} \\ q & -a_m^* \end{pmatrix} \in \Gamma_0(N) \quad (1.6)$$

is contained in the subgroup of $\Gamma_0(N)$ generated by the matrices $\sigma_1, \dots, \sigma_{2g}$.

From the analogy with CM elliptic curves this should be compared with [37, Corollary 1.3], which, as explained in the introduction of loc. cit., one should think of as an analogue of Linnik’s theorem on the least prime in an arithmetic progression (see [40, Section 2] for an in-depth discussion). We also obtain results in the case where $-1 \in H$ if we force the traces of the image of ψ to be unbalanced (see Corollaries 7.8 and 7.9).

1.3. The idea of the proof

Our proof strategy for Theorem 1.1 follows Duke’s original spectral approach as in [15]. The starting point is Weyl’s criterion (Lemma 2.4), which reduces the problem to showing that the *Weyl sums*:

$$\sum_{A \in \text{Cl}_K^+} \int_{\mathcal{C}_A} f(t) d\mu_A(t)$$

associated to an element f of an orthonormal Hecke basis of the orthogonal complement of the constant function are negligible compared to the contribution from the constant

$$\sum_{A \in \text{Cl}_K^+} |\mathcal{C}_A| = 2 |\text{Cl}_K^+| \log \epsilon_K \gg_\varepsilon D_K^{1/2-\varepsilon}, \quad (1.7)$$

where ϵ_K is the fundamental unit of K , and the ineffective lower bound is due to Siegel (via the class number formula). A slight reinterpretation of Duke's original treatment of the Weyl sums is then to relate their square to the product of L -functions

$$L\left(\frac{1}{2}, f\right) L\left(\frac{1}{2}, f, \chi_K\right),$$

where χ_K denotes the quadratic character associated to K via class field theory. This is obtained using an explicit version of Waldspurger's formula [49] (actually Duke related the Weyl sums to Fourier coefficients of half integral weight Maaß forms using an explicit correspondence of Maaß). The result now follows from a subconvexity bound for $L(1/2, f, \chi_K)$ in terms of the discriminant of K . This was obtained by Duke building on a breakthrough of Iwaniec [28].

Our approach follows the same route; on the one hand we need *lower* bounds for the contribution from the constant function, and on the other hand we need *upper* bounds for the non-constant Weyl sums. The lower bound for the constant function is achieved via a study of the small scale distribution of $(a/q, \bar{a}/q)$ on the torus $(\mathbb{R}/\mathbb{Z})^2$, where $a\bar{a} \equiv 1 \pmod{q}$; see Section 6.1. To upper bound the Weyl sums, we will use as a substitute for Waldspurger's formula the *Birch–Stevens formula* [39, Proposition 6.1], which in its original form states

$$\tau(\bar{\chi}) L(E, \chi, 1) = \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \overline{\chi(a)} 2\pi i \int_{a/q}^{i\infty} f_E(z) dz,$$

where χ is a primitive Dirichlet character modulo q , $\tau(\bar{\chi})$ is a Gauß sum and f_E is a weight 2 holomorphic form associated to an elliptic curve $E/\mathbb{Q}$. The period integrals above are known as *modular symbols*. This formula generalizes to non-primitive characters and to general Hecke–Maaß forms f by taking in place of the modular symbol, the central value of the *additive twist L-series* of f defined by analytic continuation of

$$L(s, f, x) = \sum_{n \geq 1} \frac{\lambda_f(n) e(nx)}{n^s}, \quad x \in \mathbb{Q};$$

see Proposition 5.1. The key insight in the present work (Proposition 4.2) is a relation between the geodesic integrals

$$\int_{\mathcal{C}_a^\psi(N)} f(t) d\mu_a^\psi(t)$$

and the additive twists $L(1/2, f, x)$, where $x = \psi(a)\infty$. Proving this requires a delicate analytic estimate relying on the Fourier expansion of automorphic forms and is responsible for the fact that we only get log-savings in (1.2). The result now follows by the subconvexity bound $L(1/2, f, \chi) \ll_{f,\varepsilon} q^{3/8+\varepsilon}$ for $\chi \bmod q$ due to Blomer and Harcos [3], as well as another application of the fine scale distribution of $(a/q, \bar{a}/q)$ on the torus $(\mathbb{R}/\mathbb{Z})^2$ to deal with the singular case where $(a + \bar{a} \bmod q)$ is “small”. The condition that $|H| > q^\delta$ with $\delta > 7/8$ in Theorem 1.1 comes exactly from the subconvexity exponent available, see the proof of Theorem 5.2 for more details. If one assumes the best-possible bound, the Generalized Lindelöf hypothesis, then equidistribution holds as long as $\delta > 1/2$. We however believe that any $\delta > 0$ suffices; see Conjecture 1.5 below.

In the proof of Theorem 1.3, we apply a vector space version of Weyl’s criterion (Lemma 2.6), which reduces the problem to checking a certain convergence after pairing with a basis for the dual of $H_1(Y_0(N), \mathbb{R})$. This dual is given by weight 2 holomorphic forms and the main term comes from the Eisenstein element. The contributions from holomorphic cuspidal Hecke eigenforms of weight 2 are bounded exactly as in the context of Theorem 1.1 (in fact the Weyl sums are literally the same). The lower bound for the Eisenstein contribution is achieved using an explicit version of the Birch–Stevens formula for the Eisenstein class as well as a positivity argument and lower bounds for Dirichlet L -functions at $s = 1$; see Section 6.3.

1.4. Final remarks

We will end the introduction by raising a number of natural questions in view of the real quadratic analogue (i.e. Duke’s theorem for closed geodesics).

Conjecture 1.5. *Fix $\delta > 0$. The conclusion of Theorem 1.1 holds for any subgroup $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$ such that $|H| \geq q^\delta$.*

This is a q -orbit analogue of the sparse equidistribution conjecture of Michel–Venkatesh [38, Conjecture 1] stated in the context of Heegner points. Secondly, it seems natural to try to obtain a power saving as was obtained by Duke in his original paper [15].

Conjecture 1.6. *The conclusion of Theorem 1.1 holds with a power saving error-term (in terms of q).*

Finally, it would be interesting to obtain a statement in the level aspect, meaning a version of Theorem 1.1 where the level N is allowed to vary. This was studied by Liu–Masri–Young [36] in the context of Heegner points and by Humphries and the author in [27] in the context of hyperbolic orbifolds associated to ideal classes of real quadratic fields.

2. Background

Let

$$\Gamma_0(N) := \left\{ \gamma \in \mathrm{PSL}_2(\mathbb{Z}) : \gamma \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N} \right\} \leq \mathrm{PSL}_2(\mathbb{R})$$

be the Hecke congruence group of level N considered here as a subgroup of $\mathrm{PSL}_2(\mathbb{R})$. The group $\Gamma_0(N)$ acts on $\mathbb{H} \cup \mathbf{P}^1(\mathbb{R})$ by linear fractional transformations which preserves the cusps $\mathbf{P}^1(\mathbb{Q})$. Given a matrix in $\mathrm{SL}_2(\mathbb{R})$, we will – without further mention – consider it as an element of $\mathrm{PSL}_2(\mathbb{R})$ by taking the image under the projection map.

2.1. Closed geodesics on modular surfaces

It follows from a general fact for Fuchsian groups that there is a one-to-one correspondence:

$$\{\text{oriented closed geodesics on } Y_0(N)\} \leftrightarrow \mathrm{hConj}(\Gamma_0(N)), \quad (2.1)$$

where $\mathrm{hConj}(\Gamma_0(N))$ denotes the set of hyperbolic conjugacy classes of $\Gamma_0(N)$. Explicitly, given a hyperbolic matrix

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$$

denote by $S_\gamma \subset \mathbb{H}$ the axis of γ , i.e. the fixed semi-circle of γ which has endpoints on $\mathbb{R}$ given by

$$\frac{a - d \pm \sqrt{(a + d)^2 - 4}}{2c}.$$

The bijection (2.1) is now given by mapping the conjugacy class of γ to the projection to $Y_0(N)$ of the oriented geodesic segment connecting w and γw for some $w \in S_\gamma$. Note that this projection is indeed independent of the choice of $w \in S_\gamma$ and depends only on the conjugacy class of γ . Given an oriented closed geodesic $\mathcal{C} \subset \Gamma \backslash \mathbb{H}$, we can consider the obvious lift to the unit tangent bundle

$$\mathbf{T}^1(Y_0(N)) = \Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R}),$$

recalling that $\mathbb{H} \cong \mathrm{PSL}_2(\mathbb{R})/\mathrm{PSO}_2$. By slight abuse of notation we will also denote the lift by the same symbol (e.g. $\mathcal{C}$) and will refer to these as *oriented closed geodesics* on $\mathbf{T}^1(Y_0(N))$. Let

$$a_t := \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}, \quad t \in \mathbb{R},$$

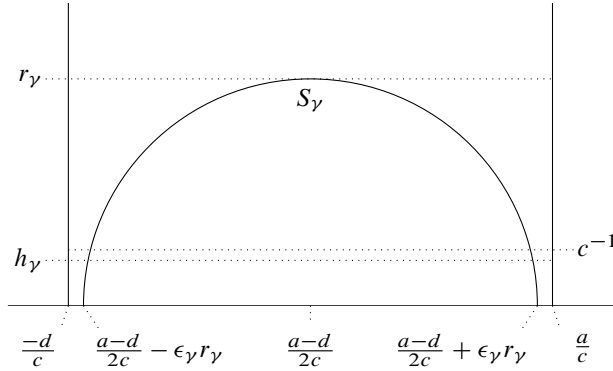


Figure 1. Key quantities associated to $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ (with the orientation corresponding to the case where $\epsilon_\gamma > 0$).

and denote by

$$A := \{a_t : t \in \mathbb{R}\} \leq \mathrm{PSL}_2(\mathbb{R}),$$

the diagonal subgroup equipped with the Haar measure dt . An oriented closed geodesics $\mathcal{C} \subset \mathbf{T}^1(Y_0(N))$ is a closed and compact A -orbit and as such the space of A -invariant measures on $\mathcal{C}$ is one-dimensional. We will denote by $\mu_{\mathcal{C}}$ the unique A -invariant measure on $\mathcal{C}$ which descends to the hyperbolic line element ds when $\mathcal{C}$ is projected to $Y_0(N)$.

Given an oriented closed geodesic $\mathcal{C} \subset Y_0(N)$ corresponding to the hyperbolic conjugacy class of $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$, we define a number of key quantities (see Figure 1 for a useful picture to have in mind). Let $\lambda_\gamma > 1$ be defined by $|\mathrm{tr}(\gamma)| = \lambda_\gamma + (\lambda_\gamma)^{-1}$ and put $\ell_\gamma := \log \lambda_\gamma$. The geodesic length of $\mathcal{C}$ then satisfies

$$|\mathcal{C}| := \int_{\mathcal{C}} ds(z) = 2\ell_\gamma \asymp \log |\mathrm{tr}(\gamma)|. \quad (2.2)$$

Furthermore, we define

$$\epsilon_\gamma := \frac{c(a+d)}{|c(a+d)|} \in \{\pm 1\}, \quad r_\gamma := \frac{\sqrt{(a+d)^2 - 4}}{2|c|} \quad (2.3)$$

(note that r_γ is the radius of S_γ) and

$$g_\gamma := (2r_\gamma)^{-1/2} \begin{pmatrix} \frac{a-d}{2c} + \epsilon_\gamma r_\gamma & \epsilon_\gamma \frac{a-d}{2c} - r_\gamma \\ 1 & \epsilon_\gamma \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{R}). \quad (2.4)$$

(Note that $c \neq 0$ since γ is hyperbolic and that the above quantities do not depend on the matrix in $\mathrm{SL}_2(\mathbb{Z})$ representing $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$). A direct calculation with matrices

shows that

$$\gamma g_\gamma a_t = g_\gamma a_{t+|\mathcal{C}|}, \quad t \in \mathbb{R},$$

which implies that we have the following standard parametrization of $\mathcal{C}$:

$$\int_{\mathcal{C}} \phi(t) d\mu_{\mathcal{C}}(t) = \int_{-|\mathcal{C}|/2}^{|\mathcal{C}|/2} \phi(g_\gamma a_t) dt = \int_{-\ell_\gamma}^{\ell_\gamma} \phi(g_\gamma a_t) dt \quad (2.5)$$

for any $\phi: \mathbf{T}^1(Y_0(N)) \rightarrow \mathbb{C}$ which is integrable with respect to $\mu_{\mathcal{C}}$. Here the left-hand side of (2.5) should be thought of as “integrating ϕ along $\mathcal{C}$ with unit speed”. Finally, we define

$$\begin{aligned} y_{\gamma,t} &:= \operatorname{Im}(g_\gamma a_t i) = \frac{2r_\gamma}{e^t + e^{-t}}, \quad t \in \mathbb{R}, \\ h_\gamma &:= y_{\gamma,\pm|\mathcal{C}|/2} = \frac{2r_\gamma}{|a+d|}. \end{aligned} \quad (2.6)$$

2.1.1. Double coset embeddings. The packets (or q -orbits) of oriented closed geodesics (1.1) that we are considering are determined by maps $\psi: (\mathbb{Z}/q\mathbb{Z})^\times \rightarrow \Gamma_0(N)$ satisfying conditions (1) and (2) above. Clearly, these conditions do not determine the packets uniquely. In this section we will characterize all possible packets of geodesics that can be obtained and prove some basic properties.

More precisely, given two positive integers N, q such that $N|q$, we will now give a complete description of the possible compositions

$$(\mathbb{Z}/q\mathbb{Z})^\times \xrightarrow{\psi} \Gamma_0(N) \xrightarrow{\pi_N} \operatorname{Conj}(\Gamma_0(N)), \quad (2.7)$$

where $\psi(a)_\infty \equiv (a/q) \bmod 1$ for all $a \in (\mathbb{Z}/q\mathbb{Z})^\times$ and $\pi_N: \Gamma_0(N) \rightarrow \operatorname{Conj}(\Gamma_0(N))$ denotes the map

$$\gamma \mapsto \{\gamma\}_{\Gamma_0(N)},$$

sending a matrix to its conjugacy class. The classical double coset decomposition (see e.g. [29, Section 2.4]) gives a bijection of sets:

$$\Gamma_\infty \backslash \Gamma_0(N) / \Gamma_\infty \leftrightarrow \left\{ \Gamma_\infty \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \Gamma_\infty \right\} \cup \bigcup_{q \geq 1: N|q} \bigcup_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \left\{ \Gamma_\infty \begin{pmatrix} a & * \\ q & * \end{pmatrix} \Gamma_\infty \right\}. \quad (2.8)$$

From this we see that the matrix $\psi(a)$ is determined by the above conditions up to left and right multiplication by Γ_∞ and we arrive at the following characterization.

Lemma 2.1. *Any composition $\pi_N \circ \psi: (\mathbb{Z}/q\mathbb{Z})^\times \rightarrow \operatorname{Conj}(\Gamma_0(N))$ as in (2.7) is a specialization of the map*

$$\Psi_N: \mathbb{Z} \oplus \Gamma_\infty \backslash \Gamma_0(N) / \Gamma_\infty \twoheadrightarrow \operatorname{Conj}(\Gamma_0(N)),$$

given by

$$\left(n, \Gamma_\infty \begin{pmatrix} a & * \\ q & * \end{pmatrix} \Gamma_\infty\right) \mapsto \left\{ \begin{pmatrix} a & b \\ q & d \end{pmatrix} \right\}_{\Gamma_0(N)},$$

where

$$ad \equiv 1 \pmod{q}, \quad \left(n - \frac{1}{2}\right)q \leq a + d < \left(n + \frac{1}{2}\right)q, \quad b = \frac{ad - 1}{q}. \quad (2.9)$$

Proof. This follows directly from the double coset decomposition (2.8) since multiplication of $\psi(a)$ by Γ_∞ on the left or right shifts the trace by multiples of q . ■

In other words, the double coset together with the choice of trace define a unique conjugacy class. Notice that all the restrictions of Ψ_N to the first component $\mathbb{Z}$ are injective (since all of the images have different traces). In our setting we will also require that the image is contained in the set of hyperbolic conjugacy classes so that they give rise to closed geodesics. Motivated by the above discussion, we arrive at the following definition.

Definition 2.2. A double coset embedding of level q is a map $\psi: (\mathbb{Z}/q\mathbb{Z})^\times \rightarrow \Gamma_0(q)$ satisfying for all $a \in (\mathbb{Z}/q\mathbb{Z})^\times$ that

- (1) $\psi(a)\infty \equiv (a/q) \pmod{1}$,
- (2) $|\mathrm{tr}(\psi(a))| > 2$.

Given a double coset embedding ψ of level q and an integer $N|q$, we denote by

$$\mathcal{C}_a^\psi(N) \subset \mathbf{T}^1(Y_0(N)), \quad (2.10)$$

the oriented closed geodesic of level N associated to the $\Gamma_0(N)$ -conjugacy class of $\psi(a) \in \Gamma_0(q) \subset \Gamma_0(N)$ under the identification (2.1). For notational simplicity we will denote the canonical measure on $\mathcal{C}_a^\psi(N)$ as defined in the previous section by

$$\mu_a^\psi := \mu_{\mathcal{C}_a^\psi(N)}, \quad (2.11)$$

suppressing the level N in the notation. For $a \in (\mathbb{Z}/q\mathbb{Z})^\times$, denote by

$$t_a := q \left(\left\{ \frac{a + \bar{a}}{q} + \frac{1}{2} \right\} - \frac{1}{2} \right) \in \mathbb{Z}, \quad (2.12)$$

the *minimal trace of a* , where $\{x\} = x - \lfloor x \rfloor$ denotes the fractional part and $\bar{a} \equiv 1 \pmod{q}$. Given a double coset embedding ψ of level q and $a \in (\mathbb{Z}/q\mathbb{Z})^\times$, we define $n_\psi(a) \in \mathbb{Z}$ by the property

$$\{\psi(a)\}_{\Gamma_0(N)} = \Psi_N \left(n_\psi(a), \Gamma_\infty \begin{pmatrix} a & * \\ q & * \end{pmatrix} \Gamma_\infty \right),$$

which makes sense by Lemma 2.1. The following lemma justifies the fact that we call t_a the minimal trace.

Lemma 2.3. *Let ψ be a double coset embedding of level q . Then for all $a \in (\mathbb{Z}/q\mathbb{Z})^\times$ it holds that*

$$\mathrm{tr}(\psi(a)) = t_a + qn_\psi(a) \quad \text{and} \quad |\mathrm{tr}(\psi(a))| \geq |t_a|. \quad (2.13)$$

Proof. We observe that $-q/2 \leq t_a < q/2$ and $t_a \equiv a + \bar{a} \pmod{q}$. Thus the first equality in (2.13) follows by the definition (2.9) of Ψ_N combined with that of $n_\psi(a)$. Furthermore, we see that indeed $|t_a|$ is the minimal possible absolute trace of $\psi(a)$ for any ψ satisfying (1), which implies the lower bound in (2.13). ■

2.2. Spectral theory of automorphic forms

For a detailed account of the following material, we refer to [29, Chapter 2] and [16, Section 4]. For a cusp $\alpha \in \mathbb{P}(\mathbb{Q})$ of $\Gamma_0(N)$, let $\Gamma_\alpha \subset \Gamma_0(N)$ be the stabilizer of α and fix a scaling matrix σ_α , i.e. any matrix satisfying

$$\sigma_\alpha^{-1} \Gamma_\alpha \sigma_\alpha = \left\{ \begin{pmatrix} 1 & \mathbb{Z} \\ & 1 \end{pmatrix} \right\}.$$

For $k \in 2\mathbb{Z}$, we denote by $\mathcal{A}(\Gamma_0(N), k)$ the vector space of all *weight k automorphic forms of level N (with trivial nebentypus)*, i.e. smooth maps $f: \mathbb{H} \rightarrow \mathbb{C}$ satisfying, for all $\gamma \in \Gamma_0(N)$ and $z \in \mathbb{H}$,

$$f(\gamma z) = j_\gamma(z)^k f(z), \quad j_\gamma(z) = \frac{j(\gamma, z)}{|j(\gamma, z)|} = \frac{cz + d}{|cz + d|}.$$

Let

$$R_k = \frac{k}{2} + (z - \bar{z}) \frac{\partial}{\partial z}, \quad \Lambda_k = \frac{k}{2} + (z - \bar{z}) \frac{\partial}{\partial \bar{z}}$$

be respectively the weight k raising and lowering operator, as defined in [16, equations (4.3)–(4.4)]. These define the maps

$$R_k: \mathcal{A}(\Gamma_0(N), k) \rightarrow \mathcal{A}(\Gamma_0(N), k+2), \quad \Lambda_k: \mathcal{A}(\Gamma_0(N), k) \rightarrow \mathcal{A}(\Gamma_0(N), k-2).$$

The weight k Laplacian is defined by

$$\begin{aligned} -\Delta_k &:= R_{k-2} \Lambda_k + \frac{k}{2} \left(1 - \frac{k}{2}\right) \\ &= \Lambda_{k+2} R_k - \frac{k}{2} \left(1 + \frac{k}{2}\right) = -y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) + ik y \frac{\partial}{\partial x}. \end{aligned}$$

For $k \in 2\mathbb{Z}_{\geq 0}$ and $s \in \mathbb{C}$, we define an operator

$$Q_{s,k}: \mathcal{A}(\Gamma_0(N), k) \rightarrow \mathcal{A}(\Gamma_0(N), k),$$

as in [16, eq. (4.65)], by

$$(Q_{s,k}f)(z) = \frac{\Gamma(s - k/2)}{\Gamma(s + k/2)} (\Lambda_{-k+2} \cdots \Lambda_{k-2} \Lambda_k f)(-\bar{z}),$$

where we put $Q_{s,k} = 0$ if $s - k/2 \in \mathbb{Z}_{\leq 0}$. Notice that for $k = 0$, we have

$$(Q_{s,0}f)(z) = f(-\bar{z}),$$

which is the usual reflection operator. The operator $Q_{s,k}$ preserves the eigenspace of $-\Delta_k$ with eigenvalue $s(1-s)$, and is an involution for $s \notin k/2 + \mathbb{Z}_{\leq 0}$. Similarly, we define $Q_{s,k}$ for negative k using the raising operators.

We say that $f \in \mathcal{A}(\Gamma_0(N), k)$ is a *Maaß form* if it satisfies:

- for all $\alpha \in \mathbb{P}(\mathbb{Q})$, $f(\sigma_\alpha z) = o(e^{2\pi y})$ as $y = \text{Im } z \rightarrow \infty$;
- f is an eigenfunction of $-\Delta_k$ with eigenvalue $\lambda_f = s_f(1-s_f) = 1/4 + t_f^2$ with $\text{Re } s_f \geq 1/2$ and $s_f = 1/2 + it_f$;
- f is an eigenfunction of the operator $Q_{s_f,k}$ with eigenvalue $\epsilon_f \in \{\pm 1, 0\}$. Note that if $k = 0$ then ϵ_f is the *sign* of the Maaß form [10, p. 106], and if f comes from a holomorphic form then $\epsilon_f = 0$.

We define the *spectral datum* of a Maaß form $f \in \mathcal{A}(\Gamma_0(N), k)$ as

$$\mathbf{t}_f := (it_f, k, \epsilon_f), \quad (2.14)$$

and we define the following convenient quantity measuring the archimedean complexity of f :

$$|\mathbf{t}_f| := |t_f| + |k| + 1.$$

Recall that the Whittaker function $W_{\alpha,\beta}: \mathbb{R}_{>0} \rightarrow \mathbb{C}$ with parameters $\alpha, \beta \in \mathbb{C}$ is the unique solution W to

$$\frac{d^2 W}{dy^2} + \left(-\frac{1}{4} + \frac{\alpha}{y} + \frac{1/4 - \beta^2}{y^2} \right) W = 0,$$

satisfying $W(y) \sim y^\alpha e^{-y/2}$ as $y \rightarrow \infty$ (with α, β fixed). In particular, we have

$$W_{0,\beta}(4\pi y) = 2y^{1/2} K_\beta(2\pi y), \quad W_{\alpha,\alpha-1/2}(4\pi y) = (4\pi y)^\alpha e^{-2\pi y}, \quad (2.15)$$

where $K_\beta(y)$ denotes the K -Bessel function. We consider the following convenient normalization of the Whittaker function:

$$\tilde{W}_{\alpha,\beta}(y) := \frac{W_{\text{sgn}(y)\alpha,\beta}(4\pi|y|)}{|\Gamma(1/2 + \beta + \text{sgn}(y)\alpha)\Gamma(1/2 - \beta + \text{sgn}(y)\alpha)|^{1/2}}, \quad y \in \mathbb{R}^\times,$$

where the right-hand side is understood as zero if $1/2 \pm \beta + \operatorname{sgn}(y)\alpha \in \mathbb{Z}_{\leq 0}$ (cf. [4, eq. (22)], where we use a trivial phase function). As explained in [9, Section 4] the functions $\tilde{W}_{\alpha,\beta}(y)$ are L^2 -normalized with respect to the measure $\frac{dy}{y}$ on $\mathbb{R}^\times$. The above conditions on the Maaß form f imply by [16, eq. 4.36)] that at any cusp $\alpha \in \mathbf{P}^1(\mathbb{Q})$, we have the Fourier expansion

$$f(\sigma_\alpha z) = f_\alpha(y) + \sum_{n \neq 0} a_{f,\alpha}(n) \tilde{W}_{k/2, it_f}(ny) e(nx), \quad (2.16)$$

$$z = x + iy, \quad e(x) = e^{2\pi i x},$$

with constant term of the shape

$$f_\alpha(y) = \begin{cases} A_\alpha y^{1/2+it_f} + B_\alpha y^{1/2-it_f}, & t_f \neq 0, \\ A_\alpha y^{1/2} + B_\alpha y^{1/2} \log y, & t_f = 0, \end{cases} \quad (2.17)$$

for some constants $A_\alpha, B_\alpha \in \mathbb{C}$. For

$$\alpha \in \mathbb{Z} \quad \text{and} \quad \beta \in i\mathbb{R} \cup \left(0, \frac{1}{2}\right) \cup \left\{\ell - \frac{1}{2} : \ell \in \mathbb{Z}_{>0}\right\},$$

we have the following uniform bounds due to Blomer–Harcos [4, eqs. (24)–(26)]:

$$\tilde{W}_{\alpha,\beta}(y) \ll |y|^{1/2} \left(\frac{|y|}{|\alpha| + |\beta| + 1} \right)^{-1-\operatorname{Re} \beta} \exp\left(-\frac{|y|}{|\alpha| + |\beta| + 1}\right), \quad (2.18)$$

and for $0 < \varepsilon < 1/4$,

$$\tilde{W}_{\alpha,\beta}(y) \ll_\varepsilon |y|^{1/2-\mathfrak{d}_\beta-\varepsilon} (|\alpha| + |\beta| + 1)^{1+\mathfrak{d}_\beta}, \quad (2.19)$$

where $\mathfrak{d}_\beta = \beta$ if $0 < \beta < 1/2$ (corresponding to the case of complementary series), and $\mathfrak{d}_\beta = 0$ otherwise. For a Maaß form f , we define

$$\mathfrak{d}_f := \mathfrak{d}_{it_f}.$$

We are interested in the spectral theory of the L^2 -space

$$\mathcal{L}^2(\Gamma_0(N), k) := \overline{\{f \in \mathcal{A}(\Gamma_0(N), k) : \langle f, f \rangle < \infty\}},$$

where

$$\langle f, g \rangle = \int_{\Gamma_0(N) \backslash \mathbb{H}} f(z) \overline{g(z)} \frac{dx dy}{y^2},$$

is the Petersson inner-product (suppressing N in the notation). We will say that f is *cuspidal at the cusp* α if $f_\alpha(y) = 0$. Finally, we say that f is a *Maaß cusp form* if it is cuspidal at all cusps $\alpha \in \mathbb{P}(\mathbb{Q})$. From [16, Corollary 4.4], we know that Maaß cusp forms arise in two ways:

- $\operatorname{Re}(s_f) < 1$ and f is obtained from repeated applications of raising or lowering operators from a weight 0 Maaß cusp form;
- $s_f = \ell/2$ with $\ell \equiv k \pmod{2}$, and f is associated, through raising or lowering operators, to a form g of weight ℓ for which $z \mapsto y^{-\ell/2}g(z)$ is holomorphic.

We will abbreviate throughout $a_f(n) = a_{f,\infty}(n)$. Moreover, by [16, eq. (4.70)] (taking into account our normalization), we have

$$\begin{aligned} a_f(-n) &= \epsilon_f \frac{|\Gamma(s_f - k/2)\Gamma(1 - s_f - k/2)|^{1/2}\Gamma(s_f + k/2)}{|\Gamma(s_f + k/2)\Gamma(1 - s_f + k/2)|^{1/2}\Gamma(s_f - k/2)} a_f(n) \\ &=: \kappa_f a_f(n). \end{aligned} \quad (2.20)$$

Note that, $\kappa_f = 0$ for f associated to a holomorphic form, whereas in the remaining cases we have $s_f \notin \mathbb{Z}$, and so by the reflection formula,

$$\begin{aligned} |\kappa_f| &= \left(\frac{|\Gamma(s_f + k/2)\Gamma(1 - s_f - k/2)|}{|\Gamma(s_f - k/2)\Gamma(1 - s_f + k/2)|} \right)^{1/2} \\ &= \left(\frac{|\pi / \sin \pi(s_f + k/2)|}{|\pi / \sin \pi(s_f - k/2)|} \right)^{1/2} = 1, \end{aligned} \quad (2.21)$$

using that $\epsilon_f \in \{\pm 1\}$ and the 2π -periodicity of sine.

2.2.1. Eisenstein series. Examples of non-cuspidal Maaß forms of even weight k include the Eisenstein series. These are not L^2 -integrable but will show up in the spectral decomposition of $\mathcal{L}^2(\Gamma_0(N), k)$. There are two natural bases for the space of Eisenstein series. The first basis is parametrized in terms of the $\Gamma_0(N)$ -equivalence classes of cusps $\alpha \in \mathbf{P}^1(\mathbb{Q})$:

$$E_{\alpha,k}(z, s) := \sum_{\gamma \in \Gamma_{\alpha} \backslash \Gamma_0(N)} j_{\sigma_{\alpha}^{-1}\gamma}(z)^{-k} \operatorname{Im}(\sigma_{\alpha}^{-1}\gamma z)^s, \quad \operatorname{Re} s > 1,$$

which satisfies meromorphic continuation with no poles on $\operatorname{Re} s = 1/2$. The other basis is in terms of characters. Let $m^2|N$ and let χ be a Dirichlet character modulo m . Then the following Eisenstein series belongs to $\mathcal{A}(\Gamma_0(N), k)$:

$$E_{\chi,k}(z, s) := \frac{1}{2} \sum_{(c,d)=1} \left(\frac{cmz + d}{|cmz + d|} \right)^{-k} \frac{(my)^s \chi(c) \chi(d)}{|cmz + d|^{2s}}, \quad s \in \mathbb{C}$$

(see [50, eq. (3.3)]), with Laplace eigenvalue $s(1-s)$ and $\mathcal{Q}_{s,k} E_{\chi,k} = \chi(-1) E_{\chi,k}$. When $\chi \pmod{m}$ is primitive and $s = 1/2 + it$, then we have the following Fourier

expansion [50, Proposition 4.1] for the Eisenstein series:

$$\begin{aligned} & \frac{i^k(m/\pi)^s}{\tau(\chi)} L(2s, \chi^2) E_{\chi,k}(z, s) \\ &= i^k \delta_{m=1} \left(\pi^{-s} \zeta(2s) y^s + \pi^{s-1} \frac{\Gamma(1-s+k/2)}{\Gamma(s+k/2)} \zeta(2-2s) y^{1-s} \right) \\ &+ \sum_{n \neq 0} \frac{\lambda_{it}(n, \chi)}{|n|^{1/2}} \frac{|\Gamma(s + \operatorname{sgn}(n)(k/2)) \Gamma(1-s + \operatorname{sgn}(n)(k/2))|^{1/2}}{\Gamma(s + \operatorname{sgn}(n)(k/2))} \\ &\quad \cdot \tilde{W}_{k/2, it}(ny) e(nx), \end{aligned}$$

where $\delta_{m=1} = 1$ if χ is the trivial character, and 0 otherwise, and

$$\lambda_{it}(n, \chi) := \chi(\operatorname{sgn}(n)) \sigma_{it, it}(|n|, \chi, \chi),$$

with

$$\begin{aligned} \sigma_{s_1, s_2}(n, \chi_1, \chi_2) &:= [(\chi_1 | \cdot |^{-s_1}) * (\overline{\chi_2} | \cdot |^{s_2})](n) \\ &= \sum_{ab=n} \chi_1(a) a^{-s_1} \overline{\chi_2(b)} b^{s_2}, \quad n \geq 1. \end{aligned} \quad (2.22)$$

It will be convenient for us to consider the following renormalization:

$$\tilde{E}_{\chi,k}(z, s) := \frac{\Gamma(s+k/2)}{|\Gamma(s+k/2)\Gamma(1-s+k/2)|^{1/2}} \frac{(m/\pi)^s}{i^k \tau(\chi)} L(2s, \chi^2) E_{\chi,k}(z, s), \quad (2.23)$$

which by the above satisfies $a_{\tilde{E}_{\chi,k}}(1) = 1$, i.e. the first Fourier coefficient is one. This will be useful when studying *additive twist L -series* of Eisenstein series in Section 3.

Consider now the weight 2 Eisenstein series $E_{\infty,2}(z, s)$ for $\operatorname{PSL}_2(\mathbb{Z})$ associated to the cusp at ∞ . Then with our normalizations we have $R_0 E_{\infty,0}(z, s) = s E_{\infty,2}(z, s)$ and $R_0 = 2iy \partial_z$, so that Hecke's limit formula [17, p. 16] amounts to

$$\lim_{s \rightarrow 0} \pi^{-s} \Gamma(s+1) \zeta(2s) E_{\infty,2}(z, s) = \frac{\pi y}{6} E_2^*(z), \quad (2.24)$$

where

$$E_2^*(z) = 1 - \frac{3}{\pi y} - 24 \sum_{n \geq 1} \sigma_1(n) e^{2\pi i n z}, \quad (2.25)$$

is the *modified Eisenstein series of weight 2*. Here $\sigma_1(n) = \sum_{d|n} d$ denotes the sum of divisors function. Note that by Hecke's limit formula $E_2^*(z)$ is a (non-holomorphic) modular form of weight 2 and level 1, meaning that

$$E_2^*(\gamma z) = j(\gamma, z)^2 E_2^*(z)$$

for $\gamma \in \operatorname{PSL}_2(\mathbb{Z})$.

2.3. Lift to the group

Consider the homogeneous space $\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R})$, which we identified with the unit tangent bundle $\mathbf{T}^1(Y_0(N))$, equipped with Haar measure dg , the Petersson inner product

$$\langle F, G \rangle = \int_{\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R})} F(g) \overline{G(g)} dg,$$

and the Laplace–Beltrami operator

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

in terms of the Iwasawa coordinates

$$g = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}.$$

We denote by $\mathcal{L}^2(\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R}), dg)$ the associated L^2 -space. For $k \in 2\mathbb{Z}$, we have a Hilbert space embedding

$$\begin{aligned} \mathcal{L}^2(\Gamma_0(N), k) &\hookrightarrow \mathcal{L}^2(\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R}), dg), \\ f &\mapsto (g \mapsto j_g(i)^{-k} f(g.i)), \end{aligned} \tag{2.26}$$

with image equal to the weight k isotypic component of $\mathcal{L}^2(\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R}), dg)$ under the (right regular) action of the maximal compact subgroup

$$K = \mathrm{PSO}_2 = \left\{ \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} : 0 \leq \theta < 2\pi \right\} / \{\pm I\},$$

i.e. for $F \in \mathcal{L}^2(\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R}), dg)$ in the image of the weight k embedding (2.26), we have

$$F \left(g \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \right) = e^{ik\theta} F(g).$$

This particular embedding intertwines the action of the weight k Laplacian $-\Delta_k$ and the Laplace–Beltrami operator Δ . We say again that the image under this embedding are of *weight k* and that the image of a Maaß form is again a *Maaß form*. We similarly extend the definition of *cuspidal*, *spectral datum* in the obvious way, and put $\mathbf{t}_F = \mathbf{t}_f$ if F is the lift of f (with notation as in (2.14)).

2.3.1. Hecke operators. For $n \geq 1$, we define the n th Hecke operator of level N as the map

$$T_n: \mathcal{C}^\infty(\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R})) \rightarrow \mathcal{C}^\infty(\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R})),$$

defined by (cf. [16, eq. (6.1)])

$$(T_n F)(g) = \frac{1}{n^{1/2}} \sum_{\substack{ad=n \\ (a,N)=1}} \sum_{0 \leq b < d} F\left(\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} g\right). \quad (2.27)$$

We say that a Maaß form $F \in \mathcal{C}^\infty(\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R}))$ is a *Hecke–Maaß form* if F is an eigenfunction for all T_n with $(n, N) = 1$. For $\ell, N' \geq 1$ such that $\ell N' | N$, we define an operator

$$v_{\ell, N'}: \mathcal{C}^\infty(\Gamma_0(N') \backslash \mathrm{PSL}_2(\mathbb{R})) \rightarrow \mathcal{C}^\infty(\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R})),$$

by

$$(v_{\ell, N'} F)(g) := F(\ell g) = F\left(\begin{pmatrix} \ell^{1/2} & 0 \\ 0 & \ell^{-1/2} \end{pmatrix} g\right), \quad g \in \mathrm{PSL}_2(\mathbb{R}). \quad (2.28)$$

We say that a Hecke–Maaß form F is a *Hecke–Maaß newform of level N* if it is not contained in the span of the images of the maps $v_{\ell, N'}$ with $N' < N$ and $\ell N' | N$ and normalized so that the first Fourier coefficient (at ∞) is 1. Note that the rescaled Eisenstein series $\tilde{E}_{\chi, k}(z, s)$ as defined in (2.23) with $\chi \bmod m$ a primitive Dirichlet character is a Hecke–Maaß newform with Hecke eigenvalues $\lambda_{it}(n, \chi)$. For $k \in 2\mathbb{Z}$, we denote the set of cuspidal newforms by

$$\mathcal{B}_k^*(\Gamma_0(N)) := \{F \in \mathcal{C}^\infty(\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R})) : \text{cuspidal Hecke–Maaß} \\ \text{newform of weight } k\}.$$

Let $F \in \mathcal{B}_k^*(\Gamma_0(N))$ be a Hecke–Maaß newform and let $L(s, \mathrm{sym}^2 F)$ denote the symmetric square L -function satisfying

$$(N + |\mathbf{t}_F|)^{-\varepsilon} \ll_\varepsilon L(1, \mathrm{sym}^2 F) \ll_\varepsilon (N + |\mathbf{t}_F|)^\varepsilon \quad (2.29)$$

for any $\varepsilon > 0$, by the classical estimates of Hoffstein–Lockhart [25, Theorem 0.2] and [16, Proposition 19.6]. It then follows from the Rankin–Selberg method [4, Section 2.1] that there exists a constant $c_N > 0$ depending only on the level N such that

$$\langle F, F \rangle = c_N L(1, \mathrm{sym}^2 F), \quad (2.30)$$

using here that $\tilde{W}_{k/2, it_f}$ is L^2 -normalized with respect to the standard Haar measure $\frac{dy}{y}$ on $\mathbb{R}^\times$ as mentioned above. The upshot is that since the level is fixed the Hecke

normalization and the L^2 -normalization are essentially equivalent (up to a factor of size $|\mathbf{t}_F|^\varepsilon$).

The spectral theorem [16, Propositions 4.1 and 4.2] states that for a smooth function $\phi: \Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R}) \rightarrow \mathbb{C}$ of compact support, we have the pointwise equality

$$\begin{aligned} \phi(g) &= \frac{\langle \phi, 1 \rangle}{\langle 1, 1 \rangle} + \sum_{F \in \mathcal{B}} \langle \phi, F \rangle F(g) \\ &\quad + \frac{1}{4\pi} \sum_{k \in 2\mathbb{Z}} \sum_{\alpha} \int_{\mathbb{R}} \langle \phi, E_{\alpha,k,it} \rangle E_{\alpha,k,it}(g) dt, \end{aligned} \quad (2.31)$$

where $\mathcal{B}$ is an orthonormal basis for the cuspidal subspace of

$$\mathcal{L}^2(\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R}), dg),$$

α runs over $\Gamma_0(N) \backslash \mathbb{P}^1(\mathbb{Q})$ (i.e. inequivalent cusps of $\Gamma_0(N)$) and

$$E_{\alpha,k,it}(g) := j_g(i)^{-k} E_{\alpha,k}\left(g.i, \frac{1}{2} + it\right),$$

with $E_{\alpha,k}$ defined as above. Here we are using that there is no non-constant residual spectrum for $\Gamma_0(N)$. It is a classical fact of Atkin–Lehner theory that the cuspidal part of the L^2 -space is spanned by cuspidal Hecke–Maaß forms and by [6, Lemma 9], we have the following orthonormal basis for the cuspidal subspace:

$$\mathcal{B}_0(\Gamma_0(N)) := \bigcup_{k \in 2\mathbb{Z}} \bigcup_{\ell N' | N} \{F_\ell : F \in \mathcal{B}_k^*(\Gamma_0(N'))\}, \quad (2.32)$$

where

$$F_\ell(g) := \sum_{d|\ell} c_F(d, \ell) F(dg), \quad (2.33)$$

with $c_F(d, \ell) \in \mathbb{C}$ satisfying $|c_F(d, \ell)| \ll_{N,\varepsilon} |\mathbf{t}_F|^\varepsilon$ (this last bound follows from the explicit formula for the basis given in [31, Proposition 2.6] and the equality (2.30) in view of the estimate (2.29)). Similarly, by combining [50, Theorem 8.1] with the last equation in Section 8.2 of loc. cit., the Eisenstein part of the spectrum has an orthonormal basis (in the sense of continuous spectrum) given by the packets

$$\mathcal{B}_e(\Gamma_0(N)) := \bigcup_{m^2 | N} \bigcup_{\substack{\chi \bmod m \\ \text{primitive}}} \bigcup_{\ell | Nm^{-2}} \{E_{\chi,k,it,\ell} : t \in \mathbb{R}\}, \quad (2.34)$$

where

$$E_{\chi,k,it,\ell}(g) := \sum_{d|\ell} c_{t,\chi}(d, \ell) j_g(i)^{-k} \tilde{E}_{\chi,k}\left(dg.i, \frac{1}{2} + it\right), \quad (2.35)$$

and the coefficients satisfy $|c_{t,\chi}(d, \ell)| \ll_{N,\varepsilon} (|t| + 1)^\varepsilon$, which follows directly from the last equation in [50, Section 8.2], the definition (2.23) of $\tilde{E}_{\chi,k}$ and the standard bound

$$|L(1 + 2it, \chi^2)| \gg_\varepsilon (|t| + 1)^{-\varepsilon}$$

(note also that $\Gamma(s + k/2)|\Gamma(s + k/2)\Gamma(1 - s + k/2)|^{-1/2}$ has absolute value 1 for $\operatorname{Re} s = 1/2$). By similar considerations, the constant term (2.17) in the Fourier expansion at ∞ of a basis element $E \in \mathcal{B}_e(\Gamma_0(N))$ satisfies for all $\varepsilon > 0$ that

$$|A_E|, |B_E| \ll_{N,\varepsilon} |\mathbf{t}_E|^\varepsilon. \quad (2.36)$$

2.4. (Co)homology of modular surfaces

Let N be prime and consider the level N modular surface $Y_0(N) = \Gamma_0(N) \backslash \mathbb{H}$. Then the compactification $X_0(N) = \overline{Y_0(N)}$ is a compact Riemann surface of genus $g = N/12 + O(1)$ (see e.g. [47, Proposition 1.40]). Recall that the integral singular homology group sits as a lattice inside the real homology (see e.g. [24, Chapter 2])

$$H_1(Y_0(N), \mathbb{Z}) \subset H_1(Y_0(N), \mathbb{R}) \cong \mathbb{R}^{2g+1},$$

which in turn sits inside the complex homology group $H_1(Y_0(N), \mathbb{C})$, and similarly for cohomology. We have the cap product pairing

$$\langle \cdot, \cdot \rangle_{\text{cap}}: H_1(Y_0(N), \mathbb{Z}) \times H^1(Y_0(N), \mathbb{Z}) \rightarrow \mathbb{Z}, \quad (2.37)$$

between integral homology and cohomology which identifies $H^1(Y_0(N), \mathbb{Z})$ with the linear dual $H_1(Y_0(N), \mathbb{Z})^*$. This extends by linearity to a pairing between real and complex (co)homology.

Given a harmonic differential 1-form on $Y_0(N)$ we get a complex cohomology class defined by integration against cycles. Explicitly, if $f \in \mathcal{M}_2(\Gamma_0(N))$ is a holomorphic form of weight 2 and level N , then

$$2\pi i f(z) dz \quad \text{and} \quad \overline{2\pi i f(z) dz}$$

are harmonic 1-forms on $Y_0(N)$. It is a consequence of the Hodge theorem that the associated cohomology classes generate $H^1(Y_0(N), \mathbb{C})$. As the Hecke operators are given in terms of correspondences on $Y_0(N)$, we get an action of the Hecke algebra (of level N) on the (co)homology groups (see [40, Section 3.2.1] for explicit formulas for the actions). A Hecke eigenbasis for $H^1(Y_0(N), \mathbb{R})$ is given by

$$B_{\text{Hecke}}(N) := \{\omega_f^\varepsilon : f \in \mathcal{B}_N, \varepsilon \in \{\pm\}\} \cup \{\omega_E(N)\}, \quad (2.38)$$

where

$$\omega_f^\pm = (2\pi i f(z) dz \pm \overline{2\pi i f(z) dz}) / (1 + i \pm \overline{1 + i}) \in H^1(Y_0(N), \mathbb{R}), \quad (2.39)$$

with $\mathcal{B}_N := \{f_1, \dots, f_g\} \subset \mathcal{S}_2(\Gamma_0(N))$ a basis of holomorphic cuspidal Hecke eigenforms of weight 2 and level N normalized so that the first Fourier coefficient is 1, and

$$\omega_E(N) := E_{2,N}(z) dz \in H^1(Y_0(N), \mathbb{R}) \quad (2.40)$$

is the normalized *Eisenstein class* defined from the Eisenstein series of weight 2 and level N :

$$E_{2,N}(z) := \frac{NE_2^*(Nz) - E_2^*(z)}{N-1} = 1 + \frac{24}{N-1}e^{2\pi iz} + \dots, \quad (2.41)$$

where E_2^* denotes the modified Eisenstein series as defined in (2.25), which is a non-holomorphic modular form of weight 2. By the Fourier expansion of E_2^* , we have that $E_{2,N}$ is holomorphic so that indeed $E_{2,N} \in \mathcal{M}_2(\Gamma_0(N))$, and we have for $(n, N) = 1$ that

$$T_n \omega_E(N) = \sigma_1(n) \omega_E(N) = (\Sigma_{d|n} d) \omega_E(N).$$

Since the cap product pairing is Hecke equivariant, one obtains a Hecke eigenbasis for the real homology by taking the dual basis of $B_{\text{Hecke}}(N)$ with respect to the cap product pairing (2.37):

$$\{v_f^\epsilon : f \in \mathcal{B}_N, \epsilon \in \{\pm\}\} \cup \{v_E(N)\},$$

where

$$\langle v_f^\pm, \omega_f^\pm \rangle_{\text{cap}} = \langle v_E(N), \omega_E(N) \rangle_{\text{cap}} = 1.$$

One can check by hand that, in fact, the dual of the Eisenstein element is given by

$$v_E(N) = [\text{class of } \{t + i : 0 \leq t < 1\}] \in H_1(Y_0(N), \mathbb{Z}),$$

since the constant term of $E_{2,N}$ is 1. Notice that geometrically $v_E(N)$ is the class of a simple loop around the cusp at ∞ and is equal to the image of $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ inside the abelianization $\Gamma_0(N)^{\text{ab}}$ modulo torsion under the isomorphism (1.4).

2.5. Weyl's criterion

The abstract setting in Theorem 1.1 is as follows: we have a sequence of probability measures $\mu_1, \mu_2, \dots$ on $\mathbf{T}^1(Y_0(N)) = \Gamma_0(N) \backslash \text{PSL}_2(\mathbb{R})$, which we want to prove converge in the weak* topology to the Haar measure dg . This means that for a smooth and compactly supported function $\phi: \mathbf{T}^1(Y_0(N)) \rightarrow \mathbb{C}$, we want to prove that

$$\int_{\mathbf{T}^1(Y_0(N))} \phi(g) d\mu_n(g) \rightarrow \int_{\mathbf{T}^1(Y_0(N))} \phi(g) dg \quad \text{as } n \rightarrow \infty.$$

By the spectral theorem (2.31) and the remarks following it, one obtains Weyl's criterion.

Lemma 2.4 (Weyl's criterion). *Let $N \geq 1$ be a fixed integer and let $\mu_1, \mu_2, \dots$ be a sequence of probability measures on $\mathbf{T}^1(Y_0(N))$. Let $A \geq 1$ be a constant and $\Phi: \mathbb{Z}_{>0} \rightarrow \mathbb{R}_{\geq 0}$ a function satisfying $\Phi(n) \rightarrow 0$ as $n \rightarrow \infty$. Assume that the following is satisfied:*

$$\left| \int_{\mathbf{T}^1(Y_0(N))} F(g) d\mu_n(g) \right| \leq |t_F|^A \Phi(n)$$

for $F \in \mathcal{B}_0(\Gamma_0(N))$, and

$$\left| \int_{\mathbf{T}^1(Y_0(N))} E(g) d\mu_n(g) \right| \leq |t_E|^A \Phi(n)$$

for $E \in \mathcal{B}_e(\Gamma_0(N))$. Here t_F (resp. t_E) denotes the spectral datum for the Hecke–Maaß cusp form F (resp. the Eisenstein series E) as defined in (2.14). Then for any smooth and compactly supported function $\phi: \mathbf{T}^1(Y_0(N)) \rightarrow \mathbb{C}$ and $\varepsilon > 0$, we have

$$\int_{\mathbf{T}^1(Y_0(N))} \phi(g) d\mu_n(g) = \int_{\mathbf{T}^1(Y_0(N))} \phi(g) dg + O_{A,\phi,\varepsilon}(\Phi(n)^{1-\varepsilon}) \quad \text{as } n \rightarrow \infty.$$

Proof. By Fourier expansion on the maximal compact, we get

$$\int_{\mathbf{T}^1(Y_0(N))} \phi(g) d\mu_n(g) = \sum_{k \in 2\mathbb{Z}} \int_{\mathbf{T}^1(Y_0(N))} \phi_k(g) d\mu_n(g),$$

in terms of $\mathrm{SO}_2(\mathbb{R})$ isotypic components:

$$\phi_k(g) := \frac{1}{2\pi} \int_0^{2\pi} \phi \left(g \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \right) e^{-(k/2)\theta i} d\theta.$$

By partial integration, we get

$$|\phi_k(g)| \ll_{m,\phi} (|k| + 1)^{-m}, \quad m \geq 1,$$

which means that we can truncate the sum at $k \ll \Phi(n)^{-\varepsilon}$ at a negligible cost (for n large enough). For the remaining terms, we spectrally expand each ϕ_k in terms of the weight k subspaces of $\mathcal{B}_0(Y_0(N)) \cup \mathcal{B}_e(Y_0(N))$ using the spectral theorem (2.31) and insert the bounds assumed for the Weyl sums. Now, for any $m \geq 1$ and $F \in \mathcal{B}_0(Y_0(N)) \cup \mathcal{B}_e(Y_0(N))$, we have

$$|\langle F, \phi_k \rangle| = |\lambda_F|^{-m} |\langle F, \Delta^m \phi_k \rangle| = |\lambda_F|^{-m} |\langle F, (\Delta^m \phi)_k \rangle| \ll_{\phi,m} (|t_F| + 1)^{-2m},$$

by self-adjointness and invariance of the Laplace–Beltrami operator. Thus we may truncate the sum at $|t_F| \ll \Phi(n)^{-\varepsilon}$ (for n large enough) at a negligible cost by the Weyl law [29, Corollary 11.2] (actually any polynomial bound on the measure of the spectrum with bounded spectral data suffices). For the remaining terms, we get the required result by the assumptions on the Weyl sums. ■

Remark 2.5. Note that complex conjugation defines an isomorphism

$$\mathcal{A}(\Gamma_0(N), k) \xrightarrow{\sim} \mathcal{A}(\Gamma_0(N), -k),$$

which implies that we may assume that the weight satisfies $k \geq 0$ when applying Weyl's criterion.

We will now derive a Weyl-type criterion in the setting of Theorem 1.3. Given a finite-dimensional real vector space V , define the (*topological*) *sphere associated to V* as

$$\mathbf{S}(V) := (V - \{0\})/\mathbb{R}_{>0},$$

equipped with the quotient topology, and denote by $v \mapsto \bar{v}$ the canonical map

$$V - \{0\} \rightarrow \mathbf{S}(V).$$

Given a sequence of elements $v_0, v_1, v_2, \dots \in V - \{0\}$, we want to give a criterion which implies that

$$\bar{v}_n \rightarrow \bar{v}_0 \quad \text{as } n \rightarrow \infty,$$

in $\mathbf{S}(V)$. Given a basis $B^* \subset V^*$ of the dual vector space V^* of V , we define the following norm on V :

$$\|v\|_{B^*} := \max_{\omega \in B^*} |\langle v, \omega \rangle_V|,$$

where $\langle \cdot, \cdot \rangle_V: V \times V^* \rightarrow \mathbb{R}$ denotes the canonical pairing. Then the corresponding Weyl criterion is given by the following.

Lemma 2.6. *Let $v_0, v_1, \dots$ be a sequence of non-zero elements in a finite-dimensional real vector space V . Let B^* be a basis for the dual V^* and $\Phi: \mathbb{Z}_{>0} \rightarrow \mathbb{R}_{\geq 0}$ a function such that*

$$|\langle v_n, \omega \rangle_V - \langle v_0, \omega \rangle_V| \leq \Phi(n) \|v_0\|_{B^*}, \quad (2.42)$$

for all $\omega \in B^*$. Then, we have

$$\left\| \frac{v_n}{\|v_n\|_{B^*}} - \frac{v_0}{\|v_0\|_{B^*}} \right\|_{B^*} \ll \Phi(n). \quad (2.43)$$

If furthermore $\Phi(n) \rightarrow 0$ as $n \rightarrow \infty$, then

$$\bar{v}_n \rightarrow \bar{v}_0 \text{ in } \mathbf{S}(V) \quad \text{as } n \rightarrow \infty. \quad (2.44)$$

Proof. The condition (2.42) implies that

$$\|v_n\|_{B^*} = \|v_0\|_{B^*} (1 + O(\Phi(n))).$$

Thus by the triangle inequality,

$$\left\| \frac{v_n}{\|v_n\|_{B^*}} - \frac{v_0}{\|v_0\|_{B^*}} \right\|_{B^*} \leq \left\| \frac{v_n - v_0}{\|v_0\|_{B^*}} \right\|_{B^*} + \|v_n\|_{B^*} \frac{|\|v_n\|_{B^*} - \|v_0\|_{B^*}|}{\|v_n\|_{B^*} \|v_0\|_{B^*}} \ll \Phi(n).$$

We have a continuous map

$$V - \{0\} \rightarrow \{v \in V : \|v\|_{B^*} = 1\} \subset V, \quad v \mapsto \frac{v}{\|v\|_{B^*}},$$

which factors through $\mathbf{S}(V)$ and the resulting map

$$\mathbf{S}(V) \xrightarrow{\sim} \{v \in V : \|v\|_{B^*} = 1\},$$

is a homeomorphism. Now the topology on V is induced by $\|\cdot\|_{B^*}$, which implies that indeed the convergence $\bar{v}_n \rightarrow \bar{v}_0$ is equivalent to the left-hand side of (2.43) converging to 0. ■

In our case, we have $V = H_1(Y_0(N), \mathbb{R})$ and identify V^* with the real cohomology $H^1(Y_0(N), \mathbb{R})$ by the cap product pairing and let $B^* = B_{\text{Hecke}}(N)$ be the Hecke basis of cohomology.

Remark 2.7. Note here that one could apply any norm on $\mathbb{R}^{\dim V}$ (in place of the sup norm) to the vector $(\langle v, \omega \rangle_V)_{\omega \in B^*}$ and a similar conclusion would still hold, since all norms on finite-dimensional vector spaces are equivalent.

3. Additive twist L -series of automorphic forms

Let $f \in \mathcal{A}(\Gamma_0(N), k)$ be a Maaß form of even weight k with Fourier expansion at ∞ :

$$f(z) = f_\infty(y) + \sum_{n \neq 0} a_f(n) \tilde{W}_{k/2, it_f}(ny) e(nx),$$

and assume that, for some $\delta > 0$, the Fourier coefficients satisfy the bound

$$\sum_{0 < |n| < X} |a_f(n)|^2 \ll_f X^{2\delta} \quad \text{as } X \rightarrow \infty. \quad (3.1)$$

In particular, this is satisfied if f is L^2 -integrable [29, Theorem 3.2] (with any $\delta > 0$) or if f is a linear combination of Eisenstein series with spectral parameter $s_f \in \mathbb{C}$, $\text{Re } s_f \geq 1/2$ (see [29, eq. (6.19)]), with $\delta = \text{Re } s_f + 1/2$. Denote by

$$F: \mathbf{T}^1(Y_0(N)) \rightarrow \mathbb{C}$$

the lift of f to the unit tangent bundle. Given $x \in \mathbb{Q}$, we define the *additive twist L -series of f* (and of F) as

$$L(s, F, x) = L(s, f, x) := \sum_{n \geq 1} \frac{a_f(n)e(nx)}{n^{s-1/2}}, \quad \operatorname{Re} s > 1 + \delta,$$

which admit meromorphic continuation to $\operatorname{Re} s > \operatorname{Re} it_f$, and for L^2 -integrable Maaß forms and for Eisenstein series with spectral parameter $s = 1/2 + it$, is regular at $s = 1/2$; see [14, Lemma 5.4]. Recall from e.g. [4, eq. (12)] that for f a Hecke–Maaß newform, we have

$$a_f(n) = \frac{\lambda_f(n)}{n^{1/2}}, \quad (3.2)$$

and so

$$L(s, f, x) = \sum_{n \geq 1} \frac{\lambda_f(n)e(nx)}{n^s}, \quad \operatorname{Re} s > 1 + \delta, \quad (3.3)$$

where $\lambda_f(n)$ denotes the n th Hecke eigenvalue of f . Notice also that for $\ell N' | N$, $f \in \mathcal{A}(\Gamma_0(N'), k)$ and $v_{\ell, N'}$ defined as in (2.28), we have the following formula for the Fourier coefficients:

$$a_{v_{\ell, N'} f}(n) = \delta_{\ell | n} a_f\left(\frac{n}{\ell}\right), \quad (3.4)$$

where $\delta_{\ell | n} = 1$ if ℓ divides n , and $\delta_{\ell | n} = 0$ otherwise, which implies that

$$L(s, v_{\ell, N'} f, x) = \ell^{1/2-s} L(s, f, \ell x). \quad (3.5)$$

This will be useful when dealing with the additive twist L -series of the bases defined in (2.32) and (2.34). We denote by

$$\gamma_{\alpha, \beta}(s) := \int_0^\infty \tilde{W}_{\alpha, \beta}(y) y^{s-1/2} \frac{dy}{y}, \quad (3.6)$$

the Mellin transform of Whittaker functions which converge absolutely for $\operatorname{Re} s > \mathfrak{d}_\beta$ by (2.18)–(2.19). Recall the definition of the constant term $f_\infty(y)$ at ∞ as in (2.17).

Proposition 3.1. *Let $\gamma \in \Gamma_0(N)$ be a hyperbolic element represented by a matrix in $\operatorname{SL}_2(\mathbb{Z})$ with lower left entry equal to $q \geq 2$. Let $f \in \mathcal{A}(\Gamma_0(N), k)$ be a Maaß form of even weight k , level N and spectral parameter*

$$it_f \in i\mathbb{R} \cup \left(0, \frac{1}{2}\right) \cup \left\{\ell - \frac{1}{2} : \ell \in \mathbb{Z}_{>0}\right\}, \quad (3.7)$$

and put

$$\gamma_s^\pm(\mathbf{t}_f) = (\kappa_f)^{(1 \mp 1)/2} q^{s-1/2} \gamma_{\pm k/2, it_f}(s),$$

with $\mathbf{t}_f$ the spectral datum of f , κ_f as defined in (2.20)² and $\gamma_{\pm k/2, it_f}(s)$ as in (3.6). Assume that f satisfies (3.1). Then, for $\operatorname{Re} s > \operatorname{Re} it_f$, we have the following equality of meromorphic functions:

$$\begin{aligned} & \int_1^\infty \left(f\left(\gamma^\infty + i\frac{y}{q}\right) - f_\infty\left(\frac{y}{q}\right) \right) y^{s-1/2} \frac{dy}{y} \\ & \quad + (-1)^{k/2} \int_1^\infty \left(f\left(\gamma^{-1}\gamma^\infty + i\frac{y}{q}\right) - f_\infty\left(\frac{y}{q}\right) \right) y^{1/2-s} \frac{dy}{y} \\ & = \gamma_s^+(\mathbf{t}_f) L(s, f, \gamma^\infty) + \gamma_s^-(\mathbf{t}_f) L(s, f, -\gamma^\infty) \\ & \quad + A_f q^{-1/2-it_f} \left(\frac{1}{s+it_f} + \frac{(-1)^{k/2}}{1-s+it_f} \right) \\ & \quad + B_f q^{-1/2+it_f} \left(\frac{1}{s-it_f} + \frac{(-1)^{k/2}}{1-s-it_f} \right), \end{aligned} \quad (3.8)$$

where in the case $t_f = 0$ the last term is replaced by

$$B_f q^{-1/2} \left(-\frac{1}{s^2} - \frac{\log q}{s} + (-1)^{k/2} \left(-\frac{1}{(s-1)^2} + \frac{\log q}{s-1} \right) \right).$$

Furthermore, the left-hand side of (3.8) defines an analytic function in $s \in \mathbb{C}$.

Proof. Note first of all that by inserting the bound (2.18) into the Fourier expansion (2.16) and using the bound for the Fourier coefficients (3.1), the left-hand side of (3.8) defines an entire function in $s \in \mathbb{C}$. By weight k automorphy of f , we have

$$\begin{aligned} f\left(\gamma^{-1}\gamma^\infty + i\frac{y}{q}\right) &= f\left(\gamma^{-1}\left(\gamma^\infty + i\frac{y^{-1}}{q}\right)\right) \\ &= j_{\gamma^{-1}}\left(\gamma^\infty + i\frac{y^{-1}}{q}\right)^k f\left(\gamma^\infty + i\frac{y^{-1}}{q}\right) \\ &= (-1)^{k/2} f\left(\gamma^\infty + i\frac{y^{-1}}{q}\right). \end{aligned}$$

Thus we conclude by the change of variable $y \mapsto 1/y$ that

$$\begin{aligned} & \int_1^\infty \left(f\left(\gamma^{-1}\gamma^\infty + i\frac{y}{q}\right) - f_\infty\left(\frac{y}{q}\right) \right) y^{1/2-s} \frac{dy}{y} \\ & = (-1)^{k/2} \int_1^\infty \left(f\left(\gamma^\infty + i\frac{y^{-1}}{q}\right) - (-1)^{k/2} f_\infty\left(\frac{y}{q}\right) \right) y^{1/2-s} \frac{dy}{y} \\ & = (-1)^{k/2} \int_0^1 \left(f\left(\gamma^\infty + i\frac{y}{q}\right) - (-1)^{k/2} f_\infty\left(\frac{y^{-1}}{q}\right) \right) y^{s-1/2} \frac{dy}{y}. \end{aligned}$$

²Here it is understood that $(\kappa_f)^{(1-1)/2} = 1$ in the case $\kappa_f = 0$.

In view of the bounds (2.18) and (2.19), (3.1) and the Fourier expansion (2.16), we conclude that for $\operatorname{Re} s \gg_f 1$ the left-hand side of (3.8) can be expressed as the sum of the following two absolutely convergent integrals:

$$\begin{aligned} & \int_0^\infty \left(f\left(\gamma\infty + i\frac{y}{q}\right) - f_\infty\left(\frac{y}{q}\right) \right) y^{s-1/2} \frac{dy}{y} \\ & \quad + \int_0^1 \left(f_\infty\left(\frac{y}{q}\right) - (-1)^{k/2} f_\infty\left(\frac{y^{-1}}{q}\right) \right) y^{s-1/2} \frac{dy}{y} \\ & = q^{s-1/2} \int_0^\infty (f(\gamma\infty + iy) - f_\infty(y)) y^{s-1/2} \frac{dy}{y} \\ & \quad + A_f q^{-1/2-it_f} \left(\frac{1}{s+it_f} + \frac{(-1)^{k/2}}{1-s+it_f} \right) \\ & \quad + B_f q^{-1/2+it_f} \left(\frac{1}{s-it_f} + \frac{(-1)^{k/2}}{1-s-it_f} \right), \end{aligned}$$

by inserting (2.17), with the last term replaced by

$$B_f q^{-1/2} \left(-\frac{1}{s^2} - \frac{\log q}{s} - (-1)^{k/2} \left(\frac{1}{(s-1)^2} - \frac{\log q}{s-1} \right) \right),$$

when $t_f = 0$. Now by inserting the Fourier expansion and interchanging sum and integral (cf. proof of [14, Lemma 5.3]), for $\operatorname{Re} s \gg_f 1$, we get

$$\begin{aligned} & \int_0^\infty (f(\gamma\infty + iy) - f_\infty(y)) y^{s-1/2} \frac{dy}{y} \\ & = \sum_{n \neq 0} a_f(n) e(n\gamma\infty) \int_0^\infty \tilde{W}_{k/2, it_f}(ny) y^{s-1/2} \frac{dy}{y} \\ & = \gamma_{k/2, it_f}(s) L(s, f, \gamma\infty) + \kappa_f \gamma_{-k/2, it_f}(s) L(s, f, -\gamma\infty), \end{aligned} \quad (3.9)$$

using that $a_f(-n) = \kappa_f a_f(n)$ by (2.20) and $\tilde{W}_{\alpha, \beta}(-y) = \tilde{W}_{-\alpha, \beta}(y)$. Now the result follows for $\operatorname{Re} s > \operatorname{Re} it_f$ by meromorphic continuation.³ ■

Remark 3.2. Note that the integral representation for the additive twist L -series in Proposition 3.1 is proved in the context of general Maaß forms for $\Gamma_0(N)$ satisfying (3.7), i.e. not restricted to elements occurring in the spectral decomposition of $\mathcal{L}^2(\Gamma_0(N), k)$. In particular, we will later apply it to the modified Eisenstein series (2.25) (a weight 2 Eisenstein series evaluated at $s = 0$).

³Note that in fact the formula (3.8) gives the meromorphic continuation of the right-hand side for all $s \in \mathbb{C}$.

Finally, we record a useful upper bound for the archimedean factors $\gamma_{\pm k/2, it_f}(s)$ at the central point $s = 1/2$.

Lemma 3.3. *Let $0 < \delta < 1/2$ and*

$$\alpha \in \mathbb{Z}, \quad \beta \in i\mathbb{R} \cup \left(0, \frac{1}{2} - \delta\right) \cup \left\{\ell - \frac{1}{2} : \ell \in \mathbb{Z}_{>0}\right\}.$$

Then we have the uniform bound

$$\gamma_{\alpha, \beta}\left(\frac{1}{2}\right) \ll_{\delta} (|\alpha| + |\beta| + 1)^{3/2}.$$

Proof. Put $M := |\alpha| + |\beta| + 1$. Then by (2.18) and (2.19), for $0 < \varepsilon < 1/4$, we have

$$\gamma_{\alpha, \beta}\left(\frac{1}{2}\right) \ll_{\varepsilon} M^{1+\mathfrak{d}_{\beta}} \int_0^M y^{1/2-\varepsilon-\mathfrak{d}_{\beta}} \frac{dy}{y} + \int_M^{\infty} y^{1/2} \exp\left(-\frac{y}{M}\right) \frac{dy}{y},$$

in all cases. If $\beta \notin (0, 1/2)$, then we get the required result by putting $\varepsilon = 1/8$, say. If $0 < \beta < 1/2 - \delta$, then we pick $\varepsilon = \delta/2$. ■

4. From geodesic periods to additive twist L -series

In this section, we will relate the integrals over closed geodesics to additive twist L -series. To motivate matters, let $f \in \mathcal{S}_2(\Gamma_0(N))$ be a holomorphic cusp form of weight 2 and level N (normalized so that the automorphy factor is $(cz + d)^2$) and denote by

$$F: \Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R}) \rightarrow \mathbb{C},$$

the lift of f to the unit tangent bundle (or, more precisely, the lift of $z \mapsto (\mathrm{Im} z)f(z)$). Let $\mathcal{C} \subset \Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R})$ be an oriented closed geodesic corresponding to the conjugacy class of $\gamma \in \Gamma_0(N)$. Then it follows from the standard parametrization (2.5) that

$$\int_{\mathcal{C}} F(t) d\mu_{\mathcal{C}}(t) = -i \int_{\mathcal{C}} f(z) dz = -i \int_{w_{\gamma}}^{\gamma w_{\gamma}} f(z) dz,$$

where w_{γ} is any point on the fixed half-circle $S_{\gamma} \subset \mathbb{H}$ of γ and the measure $\mu_{\mathcal{C}}$ is defined as in Section 2.1. Now by shifting contours (using the holomorphicity of f) and a change of variable, we have that

$$\begin{aligned} \int_{w_{\gamma}}^{\gamma w_{\gamma}} f(z) dz &= -\left(\int_{\gamma w_{\gamma}}^{\gamma w} f(z) dz + \int_{\gamma w}^w f(z) dz + \int_w^{\gamma w} f(z) dz\right) \\ &= \int_w^{\gamma w} f(z) dz \end{aligned} \tag{4.1}$$

for any $w \in \mathbb{H} \cup \mathbf{P}^1(\mathbb{Q})$. Combining with (the proof of) Proposition 3.1, we arrive at the following.

Proposition 4.1. *Let $F: \Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R}) \rightarrow \mathbb{C}$ denote the lift to the unit tangent bundle of a holomorphic cusp form $f \in \mathcal{S}_2(\Gamma_0(N))$ of weight 2. Let $\gamma \in \Gamma_0(N)$ be a hyperbolic matrix with $\gamma\infty = a/q$ with $q > 0$ and $(a, q) = 1$, and let $\mathcal{C} \subset \Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R})$ be the oriented closed geodesic corresponding to the $\Gamma_0(N)$ -conjugacy class of γ . Then it holds that*

$$\int_{\mathcal{C}} F(t) d\mu_{\mathcal{C}}(t) = -i \int_{\mathcal{C}} f(z) dz = 2L\left(\frac{1}{2}, f, \frac{a}{q}\right).$$

Proof. The proof follows by putting $w = \infty$ in (4.1), using equation (3.9) in the proof of Proposition 3.1 and recalling that $\kappa_f = 0$ and

$$\gamma_{1,1/2}\left(\frac{1}{2}\right) = \int_0^\infty \tilde{W}_{1,1/2}(y) \frac{dy}{y} = \frac{1}{\sqrt{\Gamma(2)\Gamma(1)}} \int_0^\infty 4\pi y e^{-2\pi y} \frac{dy}{y} = 2, \quad (4.2)$$

in view of (2.15). Note here that $f(z)$ is normalized so that $z \mapsto (\mathrm{Im} z) f(z)$ is a Maaß form of weight 2. ■

We would like to prove something like this for the entire automorphic spectrum of the unit tangent bundle $\Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R})$. However, in the absence of holomorphicity, we will have to settle for an approximate equality. Our main result is the following, recalling the definitions of our orthonormal bases $\mathcal{B}_0(\Gamma_0(N))$ and $\mathcal{B}_e(\Gamma_0(N))$ defined in (2.32) and (2.34).

Proposition 4.2. *Let $F \in \mathcal{B}_0(\Gamma_0(N))$ be an L^2 -normalized cuspidal Hecke–Maaß form of level N and even weight k . Let $\gamma \in \Gamma_0(N)$ be a hyperbolic matrix with $\gamma\infty = a/q$, where $q \geq 2$ and $(a, q) = 1$, and let $\mathcal{C} \subset \Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R})$ be the oriented closed geodesic corresponding to the $\Gamma_0(N)$ -conjugacy class of γ . Then, for $0 < \varepsilon < 1/4$, we have that*

$$\begin{aligned} \int_{\mathcal{C}} F(t) d\mu_{\mathcal{C}}(t) &= c_F^+ L\left(\frac{1}{2}, F, \frac{a}{q}\right) + c_F^- L\left(\frac{1}{2}, F, \frac{-a}{q}\right) \\ &\quad + O_{N,\varepsilon}\left(|\mathbf{t}_F|^{3+\varepsilon} \left(\frac{q}{|\mathrm{tr}(\gamma)|}\right)^{1/2+\varepsilon}\right), \end{aligned}$$

where $|c_F^\pm| \ll_N |\mathbf{t}_F|^{3/2}$. Similarly, for $E \in \mathcal{B}_e(\Gamma_0(N))$ an Eisenstein series, we have

$$\begin{aligned} \int_{\mathcal{C}} E(t) d\mu_{\mathcal{C}}(t) &= c_E^+ L\left(\frac{1}{2}, E, \frac{a}{q}\right) + c_E^- L\left(\frac{1}{2}, E, \frac{-a}{q}\right) \\ &\quad + O_{N,\varepsilon}\left(|\mathbf{t}_E|^{3+\varepsilon} \left(\left(\frac{q}{|\mathrm{tr}(\gamma)|}\right)^{1/2+\varepsilon} + \left(\frac{q}{|\mathrm{tr}(\gamma)|}\right)^{-1/2-\varepsilon}\right)\right), \quad (4.3) \end{aligned}$$

where $|c_E^\pm| \ll_N |\mathbf{t}_E|^{3/2}$.

The rest of the section is devoted to the proof of this result. Our approach is to use the Fourier expansion to compare the geodesic integral to the vertical line integrals which we have seen are connected to additive twist L -series. Recall from Section 2.1 the definitions of ϵ_γ , g_γ , r_γ , $y_{\gamma,t}$ given in (2.3)–(2.6) as well as the standard parametrization (2.5) of closed geodesics. The following estimates can be interpreted as saying that the closed geodesic is close to the two vertical segments as illustrated in Figure 1 and that the automorphy factors do not oscillate too much.

Lemma 4.3. *Let $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ be hyperbolic and let ℓ_γ be the hyperbolic length defined in (2.2). For $k \in 2\mathbb{Z}_{\geq 0}$, $0 < |t| < \ell_\gamma$ and $\epsilon_t = t/|t| \in \{\pm 1\}$, we have*

$$j_{g_\gamma a_t}(i)^{-k} - (-\epsilon_t)^{k/2} \ll k \frac{y_{\gamma,t}}{r_\gamma}, \quad (4.4)$$

$$\frac{-e^t + e^{-t}}{e^t + e^{-t}} + \epsilon_t \ll \frac{(y_{\gamma,t})^2}{(r_\gamma)^2}, \quad (4.5)$$

$$\mathrm{Re}(g_\gamma a_t i) - \gamma^{\epsilon_t} \infty \ll \frac{(y_{\gamma,t})^2}{r_\gamma}. \quad (4.6)$$

Proof. We start by noticing that, for $-\ell_\gamma \leq t \leq \ell_\gamma$, we have

$$e^{-|t|} \ll \frac{1}{e^t + e^{-t}} \ll \frac{y_{\gamma,t}}{r_\gamma}.$$

By direct calculation, one shows that

$$j_{g_\gamma a_t}(i) = \frac{ie^t + \epsilon_\gamma}{\sqrt{e^{2t} + 1}},$$

and thus by using the equality $x^n - y^n = (x - y)(x^{n-1} + \dots + y^{n-1})$, we arrive at

$$\begin{aligned} j_{g_\gamma a_t}(i)^{-k} - (-\epsilon_t)^{k/2} &\ll k(j_{g_\gamma a_t}(i)^{-2} + \epsilon_t) \\ &\ll k \frac{|e^{-t} - e^t + 2i\epsilon_\gamma + \epsilon_t(e^t + e^{-t})|}{e^t + e^{-t}} \ll k e^{-|t|} \ll k \frac{y_{\gamma,t}}{r_\gamma}, \end{aligned}$$

which proves (4.4). By a similar consideration, we obtain (4.5):

$$\frac{-e^t + e^{-t}}{e^t + e^{-t}} + \epsilon_t \ll \frac{|-e^t + e^{-t} + \epsilon_t(e^t + e^{-t})|}{e^t + e^{-t}} \ll e^{-2|t|} \ll \frac{(y_{\gamma,t})^2}{(r_\gamma)^2}.$$

Finally, by direct computation, we have that

$$\begin{aligned} \mathrm{Re}(g_\gamma a_t i) - \gamma^{\epsilon_t} \infty &= \frac{a-d}{2c} + \epsilon_\gamma r_\gamma \frac{e^{2t} - 1}{e^{2t} + 1} - \gamma^{\epsilon_t} \infty \\ &= \frac{a-d}{2c} + \epsilon_\gamma r_\gamma \frac{e^t - e^{-t}}{e^t + e^{-t}} - \frac{a-d + \epsilon_t(a+d)}{2c}, \end{aligned}$$

which by (4.5) is equal to

$$\begin{aligned} \epsilon_\gamma \epsilon_t r_\gamma - \frac{\epsilon_t(a+d)}{2c} + O\left(\frac{(y_{\gamma,t})^2}{r_\gamma}\right) &\ll \frac{|a+d| - \sqrt{(a+d)^2 - 4}}{2|c|} + \frac{(y_{\gamma,t})^2}{r_\gamma} \\ &\ll \frac{1}{|(a+d)c|} + \frac{(y_{\gamma,t})^2}{r_\gamma}. \end{aligned}$$

Noticing that $|a+d|^{-1} \ll (r_\gamma)^{-1}|c|^{-1}$ and $|c|^{-1} \ll y_{\gamma,t}$, we obtain

$$\frac{1}{|(a+d)c|} \ll \frac{c^{-2}}{r_\gamma} \ll \frac{(y_{\gamma,t})^2}{r_\gamma},$$

and so the final estimate (4.6) follows. $\blacksquare$

Finally, we will need the following weak but uniform estimates. Similar estimates were obtained in [14], but we will need uniformity in the spectral data similar to [4, Section 2.3].

Lemma 4.4. *Let $f \in \mathcal{A}(\Gamma_0(N), k)$ be a Hecke–Maaß form with constant term at infinity $f_\infty(y)$ and such that the lift to the unit tangent bundle lies in*

$$\mathcal{B}_0(\Gamma_0(N)) \cup \mathcal{B}_e(\Gamma_0(N)),$$

as defined in (2.32) and (2.34). Then, for $0 < \varepsilon < 1/4$, we have

$$|f(x+iy) - f_\infty(y)| \ll_{N,\varepsilon} |\mathfrak{t}_f|^{2+\varepsilon} y^{-1/2-\varepsilon}, \quad (4.7)$$

$$|f(x+iy) - f(x'+iy)| \ll_{N,\varepsilon} |\mathfrak{t}_f|^{3+\varepsilon} y^{-3/2-\varepsilon} |x-x'|. \quad (4.8)$$

Proof. Inserting the bounds (2.18) and (2.19) into the Fourier expansion (2.16), we arrive at

$$\begin{aligned} |f(x+iy) - f_\infty(y)| &= \left| \sum_{n \neq 0} a_f(n) \tilde{W}_{k/2, it_f}(ny) e(nx) \right| \\ &\ll_\varepsilon |\mathfrak{t}_f|^{1+\mathfrak{d}_f} \sum_{0 < |n| \leq |\mathfrak{t}_f|/y} |a_f(n)| (|n|y)^{1/2-\mathfrak{d}_f-\varepsilon} \\ &\quad + \sum_{|n| > |\mathfrak{t}_f|/y} |a_f(n)| (|n|y)^{1/2} \exp\left(-\frac{|n|y}{|\mathfrak{t}_f|}\right). \end{aligned}$$

By the expressions (2.33), (2.35) and the formulas (3.4), (3.2), we see that

$$|a_f(n)| \ll_{N,\varepsilon} |\mathfrak{t}_f|^\varepsilon |n|^{-1/2} \sum_{d|N} \left| \lambda_{f^*} \left(\frac{n}{(n,d)} \right) \right|,$$

where $\lambda_{f^*}(n)$ denotes the n th Hecke eigenvalue of the Hecke–Maaß newform f^* (cuspidal or Eisenstein) underlying f . Thus we conclude from [16, Proposition 19.6] the following uniform Rankin–Selberg bound:

$$\sum_{0 < |n| \leq X} |a_f(n)|^2 \ll_{N,\varepsilon} |\mathbf{t}_f|^\varepsilon X^\varepsilon \quad \text{as } X \rightarrow \infty.$$

The estimate (4.7) now follows by an application of Cauchy–Schwarz. The second claim follows by the same estimates after writing

$$f(x + iy) - f(x' + iy) = \sum_{n \neq 0} a_f(n) \tilde{W}_{k/2, it_f}(ny) (e(nx) - e(nx')),$$

and using the inequality $|e(nx) - e(nx')| \ll |n||x - x'|$. ■

With this at hand we are ready to prove our main result of this section.

Proof of Proposition 4.2. Let $F: \Gamma_0(N) \backslash \mathrm{PSL}_2(\mathbb{R}) \rightarrow \mathbb{C}$ be the lift of f , either a Hecke cusp form or an Eisenstein series as in the statement of the proposition. To lighten notation, we will simply write $r = r_\gamma$, $h = h_\gamma$, $y_t = y_{\gamma,t}$, and $\epsilon = \epsilon_t$. Using the standard parametrization (2.4) of $\mathcal{C}$, we arrive at

$$\int_{\mathcal{C}} F(g) d\mu_{\mathcal{C}}(g) = \int_{-\ell}^{\ell} F(g_\gamma a_t) dt = \int_{-\ell}^{\ell} j_{g_\gamma a_t}(i)^{-k} f(g_\gamma a_t i) dt,$$

where $2\ell = 2\ell_\gamma = |\mathcal{C}|$ is the geodesic length of $\mathcal{C}$ and g_γ as in (2.4). In view of Proposition 3.1, we want to compare this geodesic integral to the line integral of f . By a change of variable, we have for y_t as above the following equality of measures:

$$\frac{dy_t}{y_t} = \frac{-e^t + e^{-t}}{e^t + e^{-t}} dt, \quad (4.9)$$

and in view of Figure 1, we see that y_t maps both intervals $(0, \ell)$ and $(-\ell, 0)$ to the interval (h, r) . The idea is now to split up the geodesic integral according to whether $t \in (-\ell, \ell)$ is positive or negative. We will show that each of these two contributions are equal to one of the left-hand side of (3.8) in Proposition 3.1 up to a small error. More precisely, consider

$$\int_0^{\epsilon\ell} \left(j_{g_\gamma a_t}(i)^{-k} f(g_\gamma a_t i) - (-\epsilon)^{k/2+1} f(\gamma^\epsilon \infty + iy_t) \frac{-e^t + e^{-t}}{e^t + e^{-t}} \right) dt$$

for $\epsilon \in \{\pm 1\}$ (recall that $y_t = \mathrm{Im}(g_\gamma a_t i)$). We express the above as a sum of the following three terms:

$$(-\epsilon)^{k/2+1} \int_0^{\epsilon\ell} \left(f(g_\gamma a_t i) - f(\gamma^\epsilon \infty + i \mathrm{Im}(g_\gamma a_t i)) \right) \frac{-e^t + e^{-t}}{e^t + e^{-t}} dt, \quad (4.10)$$

$$\int_0^{\epsilon\ell} (f(g_\gamma a_t i) - f_\infty(y_t)) \left(j_{g_\gamma a_t}(i)^{-k} - (-\epsilon)^{k/2+1} \frac{-e^t + e^{-t}}{e^t + e^{-t}} \right) dt, \quad (4.11)$$

$$\int_0^{\epsilon\ell} f_\infty(y_t) \left(j_{g_\gamma a_t}(i)^{-k} - (-\epsilon)^{k/2+1} \frac{-e^t + e^{-t}}{e^t + e^{-t}} \right) dt. \quad (4.12)$$

Note that the last term vanishes if f is cuspidal. Applying the bounds (4.8) and (4.6) and doing the change of variable to y_t (corresponding to (4.9)), we see that (4.10) is bounded by

$$|\mathbf{t}_f|^{3+\varepsilon} \int_h^r y^{-3/2-\varepsilon} \frac{y^2}{r} \frac{dy}{y} \ll_\varepsilon |\mathbf{t}_f|^{3+\varepsilon} r^{-1/2-\varepsilon}.$$

To bound (4.11), we start by observing that for $0 \leq |t| \leq 1$ we have $y_t \asymp r$. Thus by the bound (4.7) (and bounding the remaining terms of norm 1 trivially), we get

$$\begin{aligned} \int_0^\epsilon (f(g_\gamma a_t i) - f_\infty(y_t)) \left(j_{g_\gamma a_t}(i)^{-k} - (-\epsilon)^{k/2+1} \frac{-e^t + e^{-t}}{e^t + e^{-t}} \right) dt \\ \ll_\varepsilon |\mathbf{t}_f|^{2+\varepsilon} r^{-1/2-\varepsilon}. \end{aligned}$$

To bound the remaining contribution from $(\epsilon, \epsilon\ell)$, we start by observing that by (4.4) and another application of (4.5), it holds that

$$\begin{aligned} j_{g_\gamma a_t}(i)^{-k} - (-\epsilon)^{k/2+1} \frac{-e^t + e^{-t}}{e^t + e^{-t}} &= (-\epsilon)^{k/2} - (-\epsilon)^{k/2+1} \frac{-e^t + e^{-t}}{e^t + e^{-t}} + O\left(\frac{y_t}{r}\right) \\ &\ll \frac{y_t}{r}, \end{aligned} \quad (4.13)$$

noting that $y_t \leq r$. Furthermore, for $|t| > 1$, we have $|-e^t + e^{-t}| \gg 1$, which by (4.5) gives that

$$\frac{e^t + e^{-t}}{-e^t + e^{-t}} = -\epsilon_t + O\left(\frac{y_t^2}{r^2}\right),$$

where again $\epsilon_t = t/|t|$. This implies by (4.9) that

$$dt = \frac{e^t + e^{-t}}{-e^t + e^{-t}} \frac{dy_t}{y_t} \ll \frac{dy_t}{y_t}, \quad |t| > 1$$

interpreted in the infinitesimal sense. Combining (4.13) with the bound (4.7), we see that the remaining contribution in (4.11), for t between ϵ and $\epsilon\ell$, can be bounded by

$$\int_\epsilon^{\epsilon\ell} |\mathbf{t}_f|^{2+\varepsilon} y_t^{-1/2-\varepsilon} \frac{y_t}{r} dt \ll \frac{|\mathbf{t}_f|^{2+\varepsilon}}{r} \int_h^r y^{-1/2-\varepsilon} dy \ll_\varepsilon |\mathbf{t}_f|^{2+\varepsilon} r^{-1/2-\varepsilon}.$$

Finally, in the Eisenstein case $F = E$ using the bound (2.36) for the coefficients in the constant term, the elementary inequality $|y^{1/2} \log y| \ll_\varepsilon y^{1/2+\varepsilon} + y^{1/2-\varepsilon}$ to deal

with the case where $t_E = 0$, and the considerations above, we conclude that the last term (4.12) is bounded by

$$\begin{aligned}
 &\ll \int_0^\epsilon (|A_E| + |B_E|)(y_t^{1/2+\epsilon} + y_t^{1/2-\epsilon}) dt \\
 &\quad + \int_\epsilon^{\epsilon\ell} (|A_E| + |B_E|)(y_t^{1/2+\epsilon} + y_t^{1/2-\epsilon}) dt \\
 &\ll_{N,\epsilon} |\mathbf{t}_E|^\epsilon (r^{1/2+\epsilon} + r^{1/2-\epsilon}) + |\mathbf{t}_E|^\epsilon \int_h^r (y^{1/2+\epsilon} + y^{1/2-\epsilon}) \frac{dy}{y} \\
 &\ll |\mathbf{t}_E|^\epsilon (r^{1/2+\epsilon} + r^{1/2-\epsilon}).
 \end{aligned}$$

Putting everything together and changing the coordinates back to y_t , we conclude that

$$\begin{aligned}
 &\int_{\mathcal{C}} F(g) d\mu_{\mathcal{C}}(g) \\
 &= (-1)^{k/2+1} \left(\int_h^r f(\gamma\infty + iy) \frac{dy}{y} + (-1)^{k/2} \int_h^r f(\gamma^{-1}\infty + iy) \frac{dy}{y} \right) \\
 &\quad + O_{N,\epsilon}(|\mathbf{t}_f|^{3+\epsilon}(r^{1/2+\epsilon} + r^{-1/2-\epsilon})), \tag{4.14}
 \end{aligned}$$

with the improved error-term $O_{N,\epsilon}(|\mathbf{t}_f|^{3+\epsilon}r^{-1/2-\epsilon})$ in the cuspidal case. Finally, we need to add in and subtract off the contributions from the symmetric difference of the interval (q^{-1}, ∞) and (h, r) as well as subtracting the contribution of the constant term $f_\infty(y)$. These contributions can be bounded using (4.7) and (2.36) as follows:

$$\begin{aligned}
 &\int_h^{q^{-1}} (f(\gamma^{\pm 1}\infty + iy) - f_\infty(y)) \frac{dy}{y} \ll_{N,\epsilon} |\mathbf{t}_f|^{2+\epsilon} q^{3/2+\epsilon} (q^{-1} - h) \\
 &\ll |\mathbf{t}_f|^{2+\epsilon} \frac{q^{3/2+\epsilon}}{(a+d)^2 q} \\
 &\ll |\mathbf{t}_f|^{2+\epsilon} r^{-1/2-\epsilon}, \\
 &\int_r^\infty (f(\gamma^{\pm 1}\infty + iy) - f_\infty(y)) \frac{dy}{y} \ll_{N,\epsilon} |\mathbf{t}_f|^{2+\epsilon} \int_r^\infty y^{-1/2-\epsilon} \frac{dy}{y} \\
 &\ll |\mathbf{t}_f|^{2+\epsilon} r^{-1/2-\epsilon}, \\
 &\int_h^r f_\infty(y) \frac{dy}{y} \ll_{N,\epsilon} |\mathbf{t}_f|^\epsilon (r^{1/2+\epsilon} + r^{1/2-\epsilon}).
 \end{aligned}$$

Thus, in the cuspidal case, we conclude that

$$\int_h^r f(\gamma^{\pm 1}\infty + iy) \frac{dy}{y} = \int_{q^{-1}}^\infty f(\gamma^{\pm 1}\infty + iy) \frac{dy}{y} + O_{N,\epsilon}(|\mathbf{t}_f|^{2+\epsilon}r^{-1/2-\epsilon}).$$

We substitute the above into (4.14) and apply Proposition 3.1 to get the desired formula using that

$$r = \frac{\sqrt{(a+d)^2 - 4}}{2q} \asymp \frac{|a+d|}{q}.$$

The bounds for the coefficients $c_F^\pm$ follow by combining Lemma 3.3 and the fact that $|\kappa_f| \in \{0, 1\}$; see (2.21). In the Eisenstein case, we pick up an extra error term $O_{N,\varepsilon}(|\mathbf{t}_f|^{3+\varepsilon} r^{1/2+\varepsilon})$, and then the result follows similarly by applying Proposition 3.1 and observing that the last two terms in (3.8) are bounded by

$$\ll_\varepsilon |\mathbf{t}_E|^\varepsilon q^{-1/2+\varepsilon} \ll |\mathbf{t}_E|^\varepsilon (r^{1/2+\varepsilon} + r^{-1/2-\varepsilon}),$$

by using (2.36), so that it can be absorbed into the error-term. $\blacksquare$

Remark 4.5. The proof (and the statement) of Proposition 4.2 is closely related to the fact that central values of additive twist L -series associated to Maaß forms define *quantum modular forms* in the sense of Zagier [52] as proved by Sary Drapeau and the author in [14]. These properties do not rely on the arithmetic properties of $\Gamma_0(N)$ and the corresponding statements of Proposition 4.2 for any discrete, cofinite subgroup of $\mathrm{PSL}_2(\mathbb{R})$ with a cusp at ∞ can be obtained by the same proof with a possible worse dependence on the spectral data.

5. Weyl sums

In this section we will bound the Weyl sums not contributing to the main term (i.e. the non-residual spectrum). The idea is to relate the Weyl sums to Dirichlet twists of the L -functions of the relevant Hecke–Maaß form. This is exactly the content of the Birch–Stevens formula (which takes the place of the Waldspurger formula in the automorphic approach to Duke’s theorem). First of all we will develop such a formula in sufficient generality for our purposes.

5.1. Birch–Stevens formula

As it does not require any extra work, we will for future reference in this section work with Hecke–Maaß newforms (cuspidal or Eisenstein) of *general nebentypus* $\eta_0 \bmod N$ (see e.g. [16, Section 6] for a definition). We define the additive twist L -series of such Hecke–Maaß newforms by the exact same formula (3.3) in terms of the Hecke eigenvalues.

Given a Dirichlet character $\chi \bmod q$ and a Hecke–Maaß newform f of nebentypus η_0 , which is either a cusp form or a linear combination of Eisenstein series,

we define

$$L(s, f, \chi) := \sum_{n \geq 1} \frac{\lambda_f(n) \chi(n)}{n^s}, \quad \operatorname{Re} s > 1 + \operatorname{Re}(it_f),$$

and elsewhere by meromorphic continuation, where $\lambda_f(n)$ denotes the Hecke eigenvalues of f . Note that the case of Eisenstein series (possibly off the critical line) corresponds to

$$\lambda_f(n) = \sigma_{s_f-1/2, s_f-1/2}(n, \chi_1, \chi_2),$$

where $\sigma_{s_1, s_2}(n, \chi_1, \chi_2)$ is defined in equation (2.22), $\chi_i \bmod m_i$ are primitive with $\chi_1 \chi_2(-1) = 1$, and $s_f \in \mathbb{C}$. We emphasize that the only input needed in the proof of the Birch–Stevens formula is the Hecke relations. In the case of non-primitive characters and oldforms, we will have certain local factors showing up. Given a Hecke–Maaß newform f , a Dirichlet character $\chi \bmod q$, and integers $n, \ell \geq 1$, they are defined by

$$\begin{aligned} v_s(f, \chi, n; \ell) := & \sum_{\substack{n_1 n_2 n_3 = n, \\ (n_2 n_3, \ell) | n_3}} \mu(n_1) \overline{\chi(n_1)} n_2^{1-2s} \eta_0(n_2) \mu(n_2) \chi(n_2) n_3^{1-s} \lambda_f\left(\frac{n_3}{(n_3, \ell)}\right) \\ & \times (n_3, \ell)^s \chi\left(\frac{\ell}{(n_3, \ell)}\right). \end{aligned} \quad (5.1)$$

For brevity, we put $v(f, \chi, n; \ell) := v_{1/2}(f, \chi, n; \ell)$. Then we have the following general formula.

Proposition 5.1 (Birch–Stevens formula). *Let $f \in \mathcal{A}_k(\Gamma_0(N), \eta_0)$ be a Hecke–Maaß newform of level N and nebentypus η_0 , which is either a cusp form or a linear combination of Eisenstein series. Let $\chi \bmod q$ be a Dirichlet character and $\ell \geq 1$. Then, for $s \in \mathbb{C}$, we have the following equality of meromorphic functions:*

$$\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \overline{\chi(a)} L\left(s, f, \frac{\ell a}{q}\right) = v_s\left(f, \chi^*, \frac{q}{\mathbf{c}(\chi)}; \ell\right) \tau(\bar{\chi}^*) L(s, f, \chi^*),$$

where $\chi^* \bmod \mathbf{c}(\chi)$ denotes the unique primitive character that induces χ , $\tau(\bar{\chi}^*)$ is a Gauß sum and $v_s(\dots)$ is as in (5.1).

Proof. For $\operatorname{Re} s \gg_f 1$, we have by absolute convergence that

$$\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \bar{\chi}(a) L\left(s, f, \frac{\ell a}{q}\right) = \sum_{n \geq 1} \frac{\lambda_f(n)}{n^s} \left(\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \bar{\chi}(a) e\left(\frac{\ell n a}{q}\right) \right). \quad (5.2)$$

The inner sum is a Gauß sum and by [46, Lemma 3], we get

$$\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \bar{\chi}(a) e\left(\frac{\ell n a}{q}\right) = \tau(\bar{\chi}^*) \sum_{d | (\ell n, q^*)} d \bar{\chi}^*\left(\frac{q^*}{d}\right) \mu\left(\frac{q^*}{d}\right) \chi^*\left(\frac{\ell n}{d}\right),$$

where $\chi^* \bmod \mathbf{c}(\chi)$ denotes the unique primitive character that induces χ and $q^* = q/\mathbf{c}(\chi)$. Plugging this into (5.2) and interchanging the sums, we arrive at

$$\tau(\bar{\chi}^*) \sum_{d|q^*} d \bar{\chi}^*\left(\frac{q^*}{d}\right) \mu\left(\frac{q^*}{d}\right) \sum_{\substack{n \geq 1: \\ d|\ell n}} \frac{\lambda_f(n)}{n^s} \chi^*\left(\frac{\ell n}{d}\right).$$

Now the condition $d|\ell n$ means exactly that $n = n' \frac{d}{(d, \ell)}$ for some $n' \geq 1$. Since f is a newform, we have the Hecke relations

$$\lambda_f(m_1 m_2) = \sum_{h|(m_1, m_2)} \eta_0(h) \mu(h) \lambda_f\left(\frac{m_1}{h}\right) \lambda_f\left(\frac{m_2}{h}\right),$$

which we apply with $m_1 = n'$ and $m_2 = \frac{d}{(d, \ell)}$. Thus, summing over $h|\frac{d}{(d, \ell)}$ and writing $n' = mh$ for some $m \geq 1$, we arrive at

$$\begin{aligned} & \tau(\bar{\chi}^*) \sum_{d|q^*} d \bar{\chi}^*\left(\frac{q^*}{d}\right) \mu\left(\frac{q^*}{d}\right) \sum_{h|\frac{d}{(d, \ell)}} \eta_0(h) \mu(h) \lambda_f\left(\frac{d}{h(d, \ell)}\right) \\ & \times \sum_{m \geq 1} \frac{\lambda_f(m)}{\left(\frac{d}{(d, \ell)} mh\right)^s} \chi^*\left(\frac{\ell \frac{d}{(d, \ell)} mh}{d}\right). \end{aligned}$$

Putting $n_1 = q^*/d$, $n_2 = h$, $n_3 = d/h$, using multiplicativity of χ^* , as well as the fact that

$$(n_2 n_3, \ell) | n_3 \Rightarrow (n_2 n_3, \ell) = (n_3, \ell),$$

we arrive at the required result for $\operatorname{Re} s$ sufficiently large, and thus for general s by meromorphic continuation. ■

5.2. Bounds for Weyl sums

Combining all of the above we are ready to prove our bounds for the Weyl sums. Recall Definition 2.2 of a double coset embedding $\psi: (\mathbb{Z}/q\mathbb{Z})^\times \hookrightarrow \Gamma_0(q)$, the associated oriented closed geodesics (2.10) and canonical measures (2.11), as well as our choice of orthonormal Hecke bases (2.32) and (2.34).

Theorem 5.2. *Let $F \in \mathcal{B}_0(\Gamma_0(N))$ be a cuspidal Hecke–Maaß form of level N . Consider a double coset embedding ψ of level q with $N|q$, and a coset $cH \subset (\mathbb{Z}/q\mathbb{Z})^\times$ with $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$ a subgroup. Then, for any $\varepsilon > 0$, we have*

$$\begin{aligned} \sum_{a \in cH} \int_{\mathcal{C}_a^\psi(N)} F(g) d\mu_a^\psi(g) & \ll_{N, \varepsilon} |\mathbf{t}_F|^{9/2+\varepsilon} q^{7/8+\varepsilon} \\ & + |\mathbf{t}_F|^{3+\varepsilon} \sum_{a \in cH} \left(\frac{q}{|\operatorname{tr}(\psi(a))|} \right)^{1/2+\varepsilon}. \end{aligned} \quad (5.3)$$

Similarly, for an Eisenstein series $E \in \mathcal{B}_e(\Gamma_0(N))$, we have

$$\begin{aligned} \sum_{a \in cH} \int_{\mathcal{C}_a^\psi(N)} E(g) d\mu_a^\psi(g) &\ll_{N,\varepsilon} |\mathbf{t}_E|^{11/6+\varepsilon} q^{5/6+\varepsilon} \\ &+ |\mathbf{t}_E|^{3+\varepsilon} \sum_{a \in cH} \left(\left(\frac{q}{|\operatorname{tr}(\psi(a))|} \right)^{1/2+\varepsilon} + \left(\frac{q}{|\operatorname{tr}(\psi(a))|} \right)^{-1/2-\varepsilon} \right). \end{aligned} \quad (5.4)$$

Proof. Consider first of all the cuspidal case. By character orthogonality and Proposition 4.2, we get

$$\begin{aligned} \sum_{a \in cH} \int_{\mathcal{C}_a^\psi(N)} F(g) d\mu_a^\psi(g) &= \sum_{a \in cH} \left(\left(c_F^+ L\left(\frac{1}{2}, F, \frac{a}{q}\right) + c_F^- L\left(\frac{1}{2}, F, \frac{-a}{q}\right) \right) \right. \\ &\quad \left. + O_\varepsilon\left(|\mathbf{t}_F|^{3+\varepsilon} \left(\frac{q}{|\operatorname{tr}(\psi(a))|} \right)^{1/2+\varepsilon} \right) \right) \\ &= \frac{|H|}{\varphi(q)} \sum_{\chi(q): \chi|_H=1} \chi(c) \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \overline{\chi(a)} \left(c_F^+ L\left(\frac{1}{2}, F, \frac{a}{q}\right) + c_F^- L\left(\frac{1}{2}, F, \frac{-a}{q}\right) \right) \\ &\quad + O_\varepsilon\left(|\mathbf{t}_F|^{3+\varepsilon} \sum_{a \in cH} \left(\frac{q}{|\operatorname{tr}(\psi(a))|} \right)^{1/2+\varepsilon} \right). \end{aligned} \quad (5.5)$$

Let $\ell \geq 1$ be an integer and f^* a Hecke–Maaß newform of level N' with $\ell N' | N$ such that $F = (F^*)_\ell$ as in (2.33), where F^* denotes the lift of f^* to the unit tangent bundle. Using the formula (3.5) for oldform additive twist L -series we conclude by the Birch–Stevens formula as in Proposition 5.1 that the inner-sum (in the double sum) of equation (5.5) is equal to

$$(c_F^+ + \chi(-1)c_F^-) \tau(\bar{\chi}^*) \sum_{d|\ell} c_{F^*}(d, \ell) L\left(\frac{1}{2}, f^*, \chi^*\right) \nu\left(f^*, \chi^*, \frac{q}{\mathbf{c}(\chi)}; d\right), \quad (5.6)$$

where $|c_{F^*}(d, \ell)| \ll_{N,\varepsilon} |\mathbf{t}_F|^\varepsilon$ and $|c_F^\pm| \ll_N |\mathbf{t}_F|^{3/2}$ (here the term with $d|\ell$ corresponds to the contribution from the term $F^*(dg)$ in (2.33), which in view of (5.2) shows up as the last parameter $\nu(\dots; d)$). By applying the subconvexity bound of Blomer–Harcos [3, eq. (1.3)] as well as the Kim–Sarnak bound [33] for Hecke eigenvalues to bound $\nu(\dots)$ and $|\tau(\bar{\chi}^*)| = \sqrt{\mathbf{c}(\chi)}$, the quantity (5.6) is bounded by

$$\ll_{N,\varepsilon} |\mathbf{t}_F|^{9/2+\varepsilon} q^\varepsilon \mathbf{c}(\chi)^{1/2} \mathbf{c}(\chi)^{3/8} q^{39/64} \mathbf{c}(\chi)^{-39/64} \ll_{N,\varepsilon} |\mathbf{t}_F|^{9/2+\varepsilon} q^{7/8+\varepsilon}.$$

This gives the wanted since the number of Dirichlet characters $\chi \bmod q$ such that $\chi|_H = 1$ is exactly $\varphi(q)/|H|$. In the Eisenstein case, we proceed as above using the

bound (4.3) instead: By (2.35) and the Birch–Stevens formula, there is some primitive Dirichlet character $\eta \bmod m$ with $m^2 | N$ and $t \in \mathbb{R}$ such that the inner sum is equal to

$$(c_E^+ + \chi(-1)c_E^-)\tau(\bar{\chi}^*) \sum_{d|\ell} c_{t,\eta}(d, \ell) L\left(\frac{1}{2}, \tilde{E}_{\eta,it}, \chi^*\right) v\left(\tilde{E}_{\eta,it}, \chi^*, \frac{q}{\mathbf{c}(\chi)}; d\right),$$

where $\tilde{E}_{\eta,it}(z) := \tilde{E}_{\eta,0}(z, 1/2 + it)$ in terms of the normalized Eisenstein series given in (2.23), and we have the bounds

$$|c_E^\pm| \ll_N |\mathbf{t}_E|^{3/2}, \quad |c_{t,\eta}(d, \ell)| \ll_{N,\varepsilon} |\mathbf{t}_E|^\varepsilon.$$

Since the Hecke eigenvalues of $\tilde{E}_{\eta,it}$ are given by the convolution product

$$(| \cdot |^{-it} \eta) * (| \cdot |^{it} \bar{\eta}),$$

we get the factorization

$$L\left(\frac{1}{2}, \tilde{E}_{\eta,it}, \chi^*\right) = L\left(\frac{1}{2} + it, \eta\chi^*\right) L\left(\frac{1}{2} - it, \bar{\eta}\chi^*\right).$$

Thus, by bounding the Gauß sum and the term $v(\dots)$ as above (using a simple bound for the divisor function), and using the Petrow–Young bound [42]:

$$L\left(\frac{1}{2} + it, \eta\chi^*\right) \ll_\varepsilon (\mathbf{c}(\chi)(|t| + 1)m)^{1/6+\varepsilon},$$

we conclude the bound (5.4). ■

In the case of weight 2 holomorphic forms we get the following improvement uniform in the level.

Corollary 5.3. *Let $f \in \mathcal{S}_2(\Gamma_0(N))$ be a cuspidal holomorphic Hecke eigenform of weight 2 and level N normalized, so that the first Fourier coefficient is 1. Let ψ be a double coset embedding of level q with $N|q$. Let $cH \subset (\mathbb{Z}/q\mathbb{Z})^\times$ be a coset with $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$ a subgroup. Then, we have*

$$\sum_{a \in cH} \int_{\mathcal{C}_a^\psi(N)} f(z) dz \ll_\varepsilon N^{1/4} q^{7/8+\varepsilon} + N^{3/4} q^{3/4+\varepsilon}.$$

Proof. We apply the same Fourier theoretic argument as in the proof of Theorem 5.2 combined with the Birch–Stevens formula (Proposition 5.1) to reduce to L -functions, using here Proposition 4.1 in place of Proposition 4.2 (so that there is no error-term). The result now follows by the Blomer–Harcos bound (cf. [3, Theorem 2])

$$\begin{aligned} L\left(\frac{1}{2}, f, \chi^*\right) &\ll_\varepsilon N^{1/4+\varepsilon} \mathbf{c}(\chi)^{3/8+\varepsilon} + N^{1/2+\varepsilon} (N, \mathbf{c}(\chi))^{1/4} \mathbf{c}(\chi)^{1/4+\varepsilon} \\ &\leq N^{1/4+\varepsilon} q^{3/8+\varepsilon} + N^{3/4+\varepsilon} q^{1/4+\varepsilon}, \end{aligned}$$

together with the bound

$$\tau(\bar{\chi}^*)\nu\left(f, \chi^*, \frac{q}{\mathbf{c}(\chi)}; 1\right) \ll_{\varepsilon} \mathbf{c}(\chi)^{1/2} \sum_{n_1 n_2 n_3 = q/\mathbf{c}(\chi)} n_3^{1/2+\varepsilon} \ll_{\varepsilon} q^{1/2+\varepsilon},$$

using the Ramanujan bound for weight 2 holomorphic forms and a divisor bound. ■

6. The main terms

In this section we will obtain lower bounds for the contribution corresponding to the main terms in Theorems 1.1 and 1.3. As such the considerations in this section play the role of Siegel's lower bound $|\mathrm{Cl}_K^+| \gg_{\varepsilon} D_K^{1/2-\varepsilon}$ in the automorphic approach to Duke's theorem, as explained in the introduction; see (1.7).

6.1. Distribution of multiplicative inverses

When lower bounding the lengths of q -orbits of closed geodesics, special care will have to be given to those $a \in (\mathbb{Z}/q\mathbb{Z})^{\times}$ for which $a + \bar{a}$ with $a\bar{a} \equiv 1 \pmod{q}$ is close to 0 modulo q , i.e. when the minimal trace $|t_a|$ is very small (with t_a defined as in (2.12)). This will require information about the fine scale distribution of $(a/q, \bar{a}/q)$ on the torus. This is a very classical topic (see, for instance, [26] and the references therein). We will below provide a proof of the exact estimates we will need. We will need power saving estimates for Kloosterman sums over cosets of subgroups $H \leq (\mathbb{Z}/q\mathbb{Z})^{\times}$ defined as follows for $c \in (\mathbb{Z}/q\mathbb{Z})^{\times}$:

$$S_{cH}(m, n; q) := \sum_{a \in cH} e\left(\frac{ma + n\bar{a}}{q}\right).$$

The first approach to bounding the above is to relate the above (via Fourier theory) to twisted Kloosterman sums.

Lemma 6.1. *Let $q \geq 2$ and m, n be integers. Assume that either $q|n$ or q is cube-free. Then, for any subgroup $H \leq (\mathbb{Z}/q\mathbb{Z})^{\times}$ and any coset $cH \subset (\mathbb{Z}/q\mathbb{Z})^{\times}$, we have*

$$|S_{cH}(m, n; q)| \leq \sigma_0(q)(m, n, q)^{1/2} q^{1/2}.$$

Proof. By Fourier expanding over Dirichlet characters modulo q , we get

$$S_{cH}(m, n; q) = \frac{|H|}{\varphi(q)} \sum_{\chi(q): \chi|_H = 1} \bar{\chi}(c) S_{\chi}(m, n; q), \quad (6.1)$$

where the sum is over Dirichlet characters modulo q with trivial restriction to H and

$$S_\chi(m, n; q) := \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \chi(a) e\left(\frac{ma + n\bar{a}}{q}\right).$$

If $n = 0$, then $S_\chi(m, 0; q)$ reduces to a Gauß sum, which implies by [30, Lemma 3.2] that

$$|S_\chi(m, 0; q)| = \left| \tau(\chi^*) \sum_{d|(m, q/q^*)} d \bar{\chi}^*\left(\frac{m}{d}\right) \mu\left(\frac{q}{q^*d}\right) \right| \leq \sigma_0(q)(m, q)^{1/2} q^{1/2},$$

where $\chi^* \bmod q^*$ denotes the unique primitive character inducing $\chi \bmod q$. Similarly, if q is cube-free, then we have the bound

$$|S_\chi(m, n; q)| \leq \sigma_0(q)(m, n, q)^{1/2} q^{1/2},$$

which follows from [34, Section 9], as pointed out in [26]. Inserting these bounds into the Fourier expansion (6.1), we get the required result. ■

We remark that a similar argument using [34, Section 9] yields the bound

$$|S_{cH}(m, n; q)| \leq \sigma_0(q)(m, n, q)^{1/2} q^{5/6},$$

for general q . However, using a simple Cauchy–Schwarz inequality yields the following stronger bound.⁴

Corollary 6.2. *Let $q \geq 2$, $cH \subset (\mathbb{Z}/q\mathbb{Z})^\times$ a coset of a subgroup $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$, and $m, n \in \mathbb{Z}$. Then, for all $\varepsilon > 0$, we have the bound*

$$|S_{cH}(m, n; q)| \ll_\varepsilon (m, n, q)^{1/4} q^{1/2+\varepsilon} |H|^{1/4}.$$

Proof. We start from the trivial equality

$$|H| S_{cH}(m, n; q) = \sum_{b \in H} \sum_{a \in cH} e\left(\frac{mab + n\bar{a}\bar{b}}{q}\right).$$

This implies by Cauchy–Schwarz that

$$\begin{aligned} |H|^2 |S_{cH}(m, n; q)|^2 &\leq |H| \left(\sum_{b \in H} \left| \sum_{a \in cH} e\left(\frac{mab + n\bar{a}\bar{b}}{q}\right) \right|^2 \right) \\ &= |H| \sum_{a_1, a_2 \in cH} S(m(a_1 - a_2), n(\bar{a}_1 - \bar{a}_2); q). \end{aligned}$$

⁴The author would like to thank Igor Shparlinski for suggesting this.

By the Salie–Weil bound [34, eq. (9.6)], we see that the above is bounded by

$$\leq |H| \sum_{a_1, a_2 \in cH} \sigma_0(q) q^{1/2} (m(a_1 - a_2), n(\bar{a}_1 - \bar{a}_2), q)^{1/2}.$$

Now observe that $\bar{a}_1 - \bar{a}_2 = \bar{a}_1 \bar{a}_2 (a_2 - a_1)$, which implies that

$$\begin{aligned} (m(a_1 - a_2), n(\bar{a}_1 - \bar{a}_2), q) &\leq ((a_1 - a_2)m, (a_1 - a_2)n, q) \\ &\leq (a_1 - a_2, q)(m, n, q) = (c^{-1}(a_1 - a_2), q)(m, n, q). \end{aligned}$$

Thus, we arrive at the following bound:

$$|H| |S_{cH}(m, n; q)|^2 \leq \sigma_0(q) q^{1/2} (m, n, q)^{1/2} \sum_{a_1, a_2 \in H} (a_1 - a_2, q)^{1/2}.$$

We can bound the inner sum by

$$\sum_{a_1, a_2 \in H} (a_1 - a_2, q)^{1/2} \leq \sum_{d|q} d^{1/2} \sum_{a \in H} |H \cap ((a + d\mathbb{Z})/q\mathbb{Z})|,$$

and using that $|H \cap ((a + d\mathbb{Z})/q\mathbb{Z})| \leq \min(|H|, q/d)$, we arrive at the bound

$$\begin{aligned} |H| |S_{cH}(m, n; q)|^2 &\leq \sigma_0(q) q^{1/2} (m, n, q)^{1/2} \\ &\quad \times \left(\sum_{d|q: d \leq q/|H|} d^{1/2} |H|^2 + \sum_{d|q: d > q/|H|} d^{-1/2} q |H| \right) \\ &\ll_{\varepsilon} (m, n, q)^{1/2} q^{1/2+\varepsilon} (|H|^{3/2} q^{1/2+\varepsilon}), \end{aligned}$$

which yields the required bound. ■

Note that the above yields a non-trivial bound whenever $|H| \gg q^{\delta}$ with $\delta > 2/3$.

Remark 6.3. A natural barrier for the above methods is (at best) the exponent $q^{1/2}$, which is trivial when $|H| \leq q^{1/2}$. Using different methods, Ostafe–Shparlinski–Voloch [41] obtained a power saving estimate in the case where q is prime for subgroups H of size $\gg q^{\delta}$ with $\delta > 3/7$. Most likely, one can also go beyond the square-root barrier for general prime powers using classical techniques from the theory of exponential sums. If furthermore $n = 0$, then a power saving estimate was obtained for any $\delta > 0$ by Bourgain–Konyagin [8]. We would like to thank Igor Shparlinski for pointing this out.

Now we are ready to prove the fine scale equidistribution result we will need.

Proposition 6.4. *Let $T \geq 1$ be a parameter and $\phi_T: (\mathbb{R}/\mathbb{Z})^2 \rightarrow \mathbb{C}$ a smooth function satisfying*

$$\left| \frac{\partial^n \phi_T}{\partial x^n}(x, y) \right|, \left| \frac{\partial^n \phi_T}{\partial y^n}(x, y) \right| \leq C_n T^n, \quad \forall n \geq 1,$$

for some constants $C_n > 0$ independent of T . Let $q \geq 2$ and let $cH \subset (\mathbb{Z}/q\mathbb{Z})^\times$ be a coset of a subgroup $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$. Then, for any $\varepsilon > 0$, we have

$$\sum_{a \in cH} \phi_T\left(\frac{a}{q}, \frac{\bar{a}}{q}\right) = |H| \int_{(\mathbb{R}/\mathbb{Z})^2} \phi_T(x, y) dx dy + O_\varepsilon(\|\phi_T\|_1 T^{2+\varepsilon} q^{1/2+\varepsilon} |H|^{1/4} + q^{-100}), \quad (6.2)$$

where the implied constant is allowed to depend on the constants $C_1, C_2, \dots$ and

$$\|\phi\|_1 = \int_{(\mathbb{R}/\mathbb{Z})^2} |\phi(x, y)| dx dy.$$

In the case where either q is cube-free or ϕ_T depends only on the first argument, then we have the improved error term

$$O_\varepsilon(\|\phi_T\|_1 T^{2+\varepsilon} q^{1/2+\varepsilon} + q^{-100}). \quad (6.3)$$

Proof. By Fourier expansion on the torus, we get

$$\begin{aligned} \sum_{a \in cH} \phi_T\left(\frac{a}{q}, \frac{\bar{a}}{q}\right) &= |H| \int_{(\mathbb{R}/\mathbb{Z})^2} \phi_T(x, y) dx dy \\ &+ \sum_{(m,n) \neq (0,0)} \int_{(\mathbb{R}/\mathbb{Z})^2} \phi_T(x, y) e(-mx - ny) dx dy S_{cH}(m, n; q). \end{aligned}$$

By partial integration we see that, for $(m, n) \neq (0, 0)$ with $|m| \geq |n|$ and any integer $A \geq 2$, we have by the assumption on ϕ_T that

$$\begin{aligned} \int_{(\mathbb{R}/\mathbb{Z})^2} \phi_T(x, y) e(-mx - ny) dx dy &\ll_A |m|^{-A} \int_{(\mathbb{R}/\mathbb{Z})^2} \left| \frac{\partial^A \phi_T}{\partial x^A}(x, y) \right| dx dy \\ &\ll_A \frac{T^A}{(|n| + |m|)^A}. \end{aligned}$$

Thus, we may truncate the sum at $|m|, |n| \leq q^\varepsilon T^{1+\varepsilon}$ at the cost of $O(q^{-100})$, say. By taking absolute values and applying Corollary 6.2, the remaining part of the sum is

bounded by $\|\phi_T\|_1$ multiplied by

$$\begin{aligned} & q^{1/2+\varepsilon} |H|^{1/4} \sum_{(m,n) \neq (0,0): |m|, |n| \leq q^\varepsilon T^{1+\varepsilon}} (m, n, q)^{1/4} \\ & \ll_\varepsilon q^{1/2+\varepsilon} |H|^{1/4} \sum_{d|q} d^{1/4} \left\lfloor \frac{q^\varepsilon T^{1+\varepsilon}}{d} \right\rfloor \left(\left\lfloor \frac{q^\varepsilon T^{1+\varepsilon}}{d} \right\rfloor + 1 \right) \\ & \ll_\varepsilon T^{2+\varepsilon} q^{1/2+3\varepsilon} |H|^{1/4}, \end{aligned}$$

as required. If either q is cube-free or if ϕ_T depends only on the first parameter, then we get the improved exponent by inserting the bound from Lemma 6.1 instead. This yields the estimate with the improved error term. ■

We will apply the above to obtain a number of useful estimates involving the minimal trace t_a as defined in (2.12). We recall the following standard construction of a smooth approximation of the indicator function $1_{[0,1]}: \mathbb{R} \rightarrow \{0, 1\}$ on the unit interval. Fix $\varphi: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$, a smooth bump function supported in $[-1, 1]$ with

$$\int_{-1}^1 \varphi(t) dt = 1.$$

Then, for $T \geq 1$, we define $\varphi_T(x) := T\varphi(Tx)$, which is supported in $[-T^{-1}, T^{-1}]$ and satisfies

$$\int_{-1}^1 \varphi_T(x) dx = 1.$$

This will serve as an approximation of the Dirac measure at $x = 0$, and we define the smooth function $\psi_T: \mathbb{R} \rightarrow [0, 1]$ as the following convolution:

$$\psi_T(y) = \int_{\mathbb{R}} 1_{[0,1]}(y-x) \varphi_T(x) dx,$$

satisfying the following key properties which are standard to prove:

$$|\psi_T(y)| \leq 1, \quad y \in \mathbb{R}, \quad \left| \frac{d^n \psi_T}{dy^n}(y) \right| \ll_n T^n, \quad n \geq 1, \quad (6.4)$$

$$\psi_T(y) = \begin{cases} 1, & y \in (T^{-1}, 1 - T^{-1}), \\ 0, & y \in (-\infty, -T^{-1}) \cup (1 + T^{-1}, \infty), \end{cases} \quad (6.5)$$

$$\int_{\mathbb{R}} \psi_T(x) x^m dx = \frac{1}{m+1} + O_m(T^{-1}), \quad m \geq 0.$$

Corollary 6.5. *Let $q \geq 2$ and let $cH \subset (\mathbb{Z}/q\mathbb{Z})^\times$ be a coset of a subgroup $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$. Then, for $T \geq 1$, $\delta \in (-\frac{1}{2}, \frac{1}{2})$, and $\varepsilon > 0$, the following hold:*

$$\left| \left\{ a \in cH : \left| \frac{ta}{q} - \delta \right| \leq T^{-1} \right\} \right| \ll_\varepsilon T^{-1} |H| + T^{1+\varepsilon} q^{1/2+\varepsilon} |H|^{1/4}, \quad (6.6)$$

$$\left| \left\{ a \in cH : \left| \left\{ \frac{a}{q} \right\} - \frac{1}{2} - \delta \right| \leq T^{-1} \right\} \right| \ll_\varepsilon T^{-1} |H| + T^{1+\varepsilon} q^{1/2+\varepsilon}, \quad (6.7)$$

$$\sum_{a \in cH} \frac{ta}{q} \ll_\varepsilon |H|^{3/4} q^{1/6+\varepsilon}, \quad (6.8)$$

$$\sum_{\substack{0 < a < q: \\ (a \bmod q) \in cH}} \frac{a}{q} = \frac{1}{2} |H| + O_\varepsilon(|H|^{2/3} q^{1/6+\varepsilon}). \quad (6.9)$$

Proof. For $\varphi: \mathbb{R} \rightarrow \mathbb{C}$ smooth and of compact support, we obtain a function

$$\phi: (\mathbb{R}/\mathbb{Z})^2 \rightarrow \mathbb{C}$$

by

$$\phi(x, y) := \sum_{k \in \mathbb{Z}} \varphi(x + y + k).$$

By Fubini's theorem, a change of variable, unfolding and another change of variable, we get the following useful fact:

$$\begin{aligned} \int_{(\mathbb{R}/\mathbb{Z})^2} \phi(x, y) dx dy &= \sum_{k \in \mathbb{Z}} \int_0^1 \int_k^{k+1} \varphi(x + y) dx dy \\ &= \int_0^1 \int_{-\infty}^{\infty} \varphi(x + y) dx dy = \int_{-\infty}^{\infty} \varphi(x) dx. \end{aligned} \quad (6.10)$$

The first estimate follows directly from Proposition 6.4 and (6.10) by letting

$$\phi_T(x, y) = \sum_{k \in \mathbb{Z}} \varphi(T(x + y + k - \delta)),$$

where $\varphi: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ is some smooth positive function with compact support contained in $(-1/2, 1/2)$ and $\varphi(y) = 1$ for $y \in (-1/4, 1/4)$, say, so that $\|\phi_T\|_1 \asymp T^{-1}$. The second estimate follows similarly by applying Proposition 6.4 with the improved error term (6.3) to

$$\phi_T(x, y) = \sum_{k \in \mathbb{Z}} \varphi\left(T\left(x + k - \delta - \frac{1}{2}\right)\right).$$

To prove the third estimate, note first of all that the left-hand side is trivially bounded by $(1/2)|H|$, and so we can assume that $|H| \geq q^{2/3}$. Consider now the test function

$$\phi_T(x, y) = -\frac{1}{2} + \sum_{k \in \mathbb{Z}} \left(x + y + k + \frac{1}{2}\right) \psi_T\left(x + y + k + \frac{1}{2}\right),$$

with ψ_T defined as above. Then we have by the definition (2.12) of t_a , the properties (6.4), (6.5), and the estimate (6.6) that

$$\begin{aligned} \sum_{a \in cH} \left| \frac{t_a}{q} - \phi_T \left(\frac{a}{q}, \frac{\bar{a}}{q} \right) \right| &\ll \# \left\{ a \in cH : \frac{t_a}{q} \in \left[-\frac{1}{2} - T^{-1}, -\frac{1}{2} \right] \cup \left[\frac{1}{2}, \frac{1}{2} + T^{-1} \right] \right\} \\ &\ll_{\varepsilon} T^{-1} |H| + T^{1+\varepsilon} q^{1/2+\varepsilon} |H|^{1/4}. \end{aligned}$$

Now we apply Proposition 6.4 with the test function ϕ_T . By the equality (6.10) and the estimate (6.5), we see that the main term integral amounts to

$$\int_{(\mathbb{R}/\mathbb{Z})^2} \phi_T(x, y) dx dy = -\frac{1}{2} + \int_{\mathbb{R}} \psi_T(x) x dx = -\frac{1}{2} + \frac{1}{2} + O(T^{-1}) \ll T^{-1},$$

and so using the trivial bound $\|\phi_T\|_1 \ll 1$, we arrive at the bound

$$\ll_{\varepsilon} T^{-1} |H| + T^{1+\varepsilon} q^{1/2+\varepsilon} |H|^{1/4} + T^{2+\varepsilon} q^{1/2+\varepsilon} |H|^{1/4}.$$

Picking $T = |H|^{1/4} q^{-1/6} \geq 1$ yields the required result (using here that $|H| \geq q^{2/3}$). The final estimate follows similarly using the test function

$$\phi_T(x, y) = \sum_{k \in \mathbb{Z}} (x + k) \psi_T(x + k),$$

the improved error term (6.3), the previous estimate (6.7), and picking

$$T = |H|^{1/3} q^{-1/6}$$

when $|H| \geq q^{1/2}$. ■

6.2. Lengths of closed geodesics

A key application of the above is the following lower bound for the closed geodesics associated to an arbitrary double coset embedding.

Proposition 6.6. *Let $N \geq 1$ and $\delta > 2/3$. Let $q \geq 2$ be an integer such that $N|q$ and consider a double coset embedding ψ of level q . Let $H \leq (\mathbb{Z}/q\mathbb{Z})^{\times}$ be a subgroup with $|H| \geq q^{\delta}$ and $cH \subset (\mathbb{Z}/q\mathbb{Z})^{\times}$ a coset. Then, we have*

$$\sum_{a \in cH} |\mathcal{C}_a^{\psi}(N)| \gg |H| \log q.$$

If furthermore q is cube-free, the same conclusion holds as long as $\delta > 1/2$.

Proof. We apply the estimate (6.2) (resp. (6.3)) from Proposition 6.4 (with $T = 1$) to the function

$$(x, y) \mapsto \sum_{k \in \mathbb{Z}} \varphi(x + y + k),$$

where $\varphi: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ is some smooth bump function with support contained in the interval $(1/5, 1/2)$ and $\varphi(y) = 1$ for $y \in (1/4, 1/3)$, say. This implies that for a positive proportion of $a \in cH$ we have that $|t_a| \geq (1/6)q$, with t_a defined as in (2.12). Thus, we conclude by Lemma 2.3 that, for a positive proportion of $a \in cH$, we have

$$|\mathrm{tr}(\psi(a))| \geq \left(\frac{1}{6}\right)q,$$

and thus by (2.2), it follows that $|\mathcal{C}_a^\psi(N)| \gg \log q$. This gives the required lower bound for the geodesic lengths. $\blacksquare$

Remark 6.7. Note that in the case where q is prime, a similar lower bound holds whenever $|H| \gg q^\delta$ with $\delta > 3/7$ by the results mentioned in Remark 6.3.

6.3. Eisenstein contribution in homology

In this section we will determine the main term in the homological equidistribution problem as in Theorem 1.3 (i.e. the contribution coming from the Eisenstein element). Combining Hecke's limit formula (2.24), the Fourier expansion (2.25) of the modified Eisenstein series E_2^* and the formula for the normalized Whittaker function as in (4.2), we conclude that

$$\begin{aligned} E(z) &:= \frac{-\pi y}{6} E_2^*(z) \\ &= -\frac{\pi}{6}y + \frac{1}{2} + 4\pi y \sum_{n \geq 1} \sigma_1(n) e^{2\pi i n z} \\ &= -\frac{\pi}{6}y + \frac{1}{2} + \sum_{n \geq 1} \frac{\sigma_1(n)}{n} \tilde{W}_{1,1/2}(4\pi n y) e(nx), \end{aligned} \quad (6.11)$$

is a Hecke–Maaß newform of weight 2 and level 1 (since $E_{\infty,2}(z, s)$ is of weight 2 and level 1). Furthermore, again by the limit formula (2.24) and the fact that Eisenstein series are Hecke eigenforms [50, eq. (4.13)] we conclude (by interchanging the limit and the finite sum in the definition (2.27)) that

$$T_n E = \frac{\sigma_1(n)}{n^{1/2}} E = \sigma_{1/2,1/2}(n) E, \quad n \geq 1,$$

where T_n denotes the Hecke operator as defined in (2.27),

$$\sigma_{s_1, s_2}(n) = \sum_{ab=n} a^{-s_1} b^{s_2} \quad \text{and} \quad \sigma_1(n) = \sum_{d|n} d.$$

In view of this, the local weights (5.1) are given by

$$\begin{aligned} v(\chi, q; \ell) &:= v_{1/2}(E, \chi, q; \ell) \\ &= \sum_{\substack{n_1 n_2 n_3 = q, \\ (n_2 n_3, \ell) | n_3}} \mu(n_1) \overline{\chi(n_1)} \mu(n_2) \chi(n_2) \sigma_1\left(\frac{n_3}{(n_3, \ell)}\right) (n_3, \ell) \chi\left(\frac{\ell}{(n_3, \ell)}\right), \end{aligned} \quad (6.12)$$

which, as we shall see, take a particularly nice form. The key result is the following Birch–Stevens formula for the Eisenstein class.

Proposition 6.8. *Let $N \geq 2$ be prime and denote by $\omega_E(N)$ the Eisenstein class in cohomology as defined in (2.40). For each $q \geq 2$ such that $N|q$, consider a double coset embedding ψ of level q and a Dirichlet character $\chi \bmod q$. Then, we have*

$$\begin{aligned} &\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \overline{\chi(a)} \langle [\mathfrak{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}} \\ &= \frac{6(1 - \chi(-1))}{\pi^2} \frac{|\chi^*(N) - 1|^2}{N - 1} |L(1, \chi^*)|^2 \mathfrak{c}(\chi) v\left(\chi^*, \frac{q^*}{(q^*, N^\infty)}; 1\right) \\ &\quad + \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \overline{\chi(a)} \frac{\text{tr}(\psi(a))}{q}, \end{aligned} \quad (6.13)$$

where $\chi^* \bmod \mathfrak{c}(\chi)$ is the unique primitive character inducing χ and $q^* = q/\mathfrak{c}(\chi)$. Here the first term is interpreted as 0 if χ is principal.

In proving this, we will need the following lemma on the local weights appearing in the Birch–Stevens formula as in Proposition 5.1. It will be convenient to define

$$\sigma_1(p^m) := \frac{p^{m+1} - 1}{p - 1} \quad (6.14)$$

for all primes p and integers $m \in \mathbb{Z}$.

Lemma 6.9. *Let $q \geq 2$ and let $\chi \bmod D$ be a Dirichlet character (not necessarily primitive). Then the weight $v(\chi, q; 1)$ is a real number satisfying*

$$v(\chi, q; 1) \geq \varphi(q).$$

More precisely, for p prime and $m \geq 1$, we have

$$v(\chi, p^m; 1) = \frac{p^{m-1} |p - \chi(p)|^2 - |1 - \chi(p)|^2}{p - 1}.$$

Proof. We observe that $v(\chi, q; 1)$ is a convolution of multiplicative functions, and so is itself a multiplicative function of q . By definition, we see that

$$\begin{aligned} v(\chi, p^m; 1) &= \sigma_1(p^m) - \sigma_1(p^{m-1}) (\chi(p) + \overline{\chi(p)}) \\ &\quad + \sigma_1(p^{m-2}) |\chi(p)|^2, \quad m \geq 1, \end{aligned} \quad (6.15)$$

which also holds for $m = 1$ using the above extension (6.14) of σ_1 . We conclude the required formula after a short calculation. Now it follows readily from (6.15) that the expression is minimized when $\chi(p) = 1$, in which case

$$v(\chi, p^m; 1) = p^{m-1}(p-1) = \varphi(p^m).$$

Thus the result follows by multiplicativity. $\blacksquare$

Lemma 6.10. *Let $q \geq 2$ be an integer, ℓ a prime and $\chi \bmod D$ a Dirichlet character (not necessarily primitive). Then, we have*

$$v(\chi, q; \ell) - v(\chi, q; 1) = \begin{cases} -(1 - \chi(\ell))v(\chi, q; 1), & (\ell, q) = 1, \\ -|1 - \chi(\ell)|^2 v(\chi, q/(q, \ell^\infty); 1), & \ell | q. \end{cases}$$

Proof. Let $\ell^m \parallel q$ with $m \geq 0$. We start by observing that by the explicit formula for the weights (6.12), we have the factorization

$$v(\chi, q; \ell) = v(\chi, \ell^m; \ell) v\left(\chi, \frac{q}{\ell^m}; 1\right).$$

This means that $v(\chi, q; \ell) = v(\chi, q; 1)\chi(\ell)$ if $(q, \ell) = 1$, which yields the wanted formula in this case. Now assume that $m \geq 1$. We see again by the explicit formula (6.12) that

$$v(\chi, \ell^m; \ell) = \ell \sigma_1(\ell^{m-1}) - \ell \sigma_1(\ell^{m-2})(\chi(\ell) + \bar{\chi}(\ell)) + \ell \sigma_1(\ell^{m-3})|\chi(\ell)|^2,$$

using again the extension (6.14) of σ_1 . Subtracting the formula (6.15) from the above and using that $\ell \sigma_1(\ell^{j-1}) - \sigma_1(\ell^j) = -1$ for any $j \in \mathbb{Z}$, yield the wanted formula for $m \geq 1$. $\blacksquare$

Proof of Proposition 6.8. If $\chi = \chi_0$ is principal then the first term of (6.13) vanishes. The explicit formula for integrals of the holomorphic Eisenstein series of level N along closed geodesics in terms of Dedekind sums $s(a, q)$ as in [40, eq. (3.20)] gives

$$\begin{aligned} \langle [\mathcal{E}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}} &= \frac{1}{N-1} \left[\frac{\text{tr}(\psi(a))}{q/N} - 12s\left(a, \frac{q}{N}\right) - 3 \text{sgn}(\text{tr}(\psi(a))) \right. \\ &\quad \left. - \left(\frac{\text{tr}(\psi(a))}{q} - 12s(a, q) - 3 \text{sgn}(\text{tr}(\psi(a))) \right) \right] \\ &= \frac{12(s(a, q) - s(a, q/N))}{N-1} + \frac{\text{tr}(\psi(a))}{q}. \end{aligned}$$

The result follows in this case by summing the above over $a \in (\mathbb{Z}/q\mathbb{Z})^\times$ and using the standard symmetry $s(-a, q) = -s(a, q)$.

Now let us turn to the case where χ is non-principal. We have by holomorphicity

$$\langle [\mathcal{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}} = \int_w^{\psi(a)w} E_{2,N}(z) dz \quad (6.16)$$

for any $w \in \mathbb{H}$. Let

$$x_a^+ := \psi(a)\infty, \quad x_a^- := \psi(a)^{-1}\infty,$$

and put $w = x_a^- + iu/q$ in equation (6.16), which satisfies

$$\psi(a)\left(x_a^- + i\frac{u}{q}\right) = x_a^+ + i\frac{u^{-1}}{q} \quad \text{and} \quad j_{\psi(a)}(x_a^- + iu)^2 = -1, \quad u > 0. \quad (6.17)$$

By choosing the contour consisting of two straight lines connecting w , $x_a^- + iu^{-1}/q$ and $\psi(a)w$, we arrive at

$$\begin{aligned} \langle [\mathcal{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}} &= x_a^+ - x_a^- + i \int_{u/q}^{1/q} E_{2,N}(x_a^- + iy) dy \\ &\quad + i \int_{1/q}^{u^{-1}/q} E_{2,N}(x_a^- + iy) dy. \end{aligned} \quad (6.18)$$

Recalling that $z \mapsto \text{Im}(z)E_{2,N}(z)$ is automorphic of weight 2 we conclude by the change of variable $y \mapsto y/q$, the properties (6.17) and the change of variable $y \mapsto y^{-1}/q$ that

$$\begin{aligned} \int_{u/q}^{1/q} E_{2,N}(x_a^- + iy) dy &= \int_u^1 \frac{y}{q} E_{2,N}\left(x_a^- + i\frac{y}{q}\right) \frac{dy}{y} \\ &= - \int_u^1 \frac{y^{-1}}{q} E_{2,N}\left(x_a^+ + i\frac{y^{-1}}{q}\right) \frac{dy}{y} \\ &= - \int_{1/q}^{u^{-1}/q} E_{2,N}(x_a^+ + iy) dy. \end{aligned}$$

Now we insert the above into (6.18) and subtract 1 from each integrand (which cancel out). By the Fourier expansion (2.41), $E_{2,N}(z) - 1$ decays rapidly as $\text{Im } z \rightarrow \infty$, and so by letting $u \rightarrow 0$, we conclude

$$\begin{aligned} \langle [\mathcal{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}} &= x_a^+ - x_a^- + i \lim_{u \rightarrow 0} \int_{u/q}^{u^{-1}/q} E_{2,N}(x_a^- + iy) dy \\ &= \frac{\text{tr}(\psi(a))}{q} + i \int_{1/q}^{\infty} (E_{2,N}(x_a^- + iy) - 1) dy \\ &\quad - i \int_{1/q}^{\infty} (E_{2,N}(x_a^+ + iy) - 1) dy. \end{aligned}$$

Using the definition of $E_{2,N}$, we can write the integrals in terms of the modified Eisenstein series (2.25) as

$$\int_{1/q}^{\infty} (E_{2,N}(x_a^{\pm} + iy) - 1) dy = \frac{1}{N-1} \left(\int_{1/q}^{\infty} (NE_2^*(N(x_a^{\pm} + iy)) - N + \frac{3}{\pi y}) dy \right. \\ \left. - \int_{1/q}^{\infty} (E_2^*(x_a^{\pm} + iy) - 1 + \frac{3}{\pi y}) dy \right).$$

Now we apply Proposition 3.1 (after the change of variable $y/q \mapsto y$) for $f = v_{\ell,1}E$ with $\ell \in \{1, N\}$ and E defined as in (6.11) which, for $s \in \mathbb{C}$, yields

$$\int_{1/q}^{\infty} \left(\ell E_2^*(\ell(x_a^+ + iy)) - \ell + \frac{3}{\pi y} \right) (qy)^{s-1/2} dy \\ - \int_{1/q}^{\infty} \left(\ell E_2^*(\ell(x_a^- + iy)) - \ell + \frac{3}{\pi y} \right) (qy)^{1/2-s} dy \\ = \frac{-6}{\pi} \int_{1/q}^{\infty} \left(E(\ell(x_a^+ + iy)) + \frac{\pi}{6} \ell y - \frac{1}{2} \right) (qy)^{s-1/2} \frac{dy}{y} \\ + \frac{6}{\pi} \int_{1/q}^{\infty} \left(E(\ell(x_a^- + iy)) + \frac{\pi}{6} \ell y - \frac{1}{2} \right) (qy)^{1/2-s} \frac{dy}{y} \\ = \frac{-6}{\pi} \left(q^{s-1/2} 2(2\pi\ell)^{1/2-s} \Gamma\left(s + \frac{1}{2}\right) L(s, E, \ell\gamma\infty) \right. \\ \left. + \frac{\pi\ell}{6q} \left(\frac{1}{3/2-s} - \frac{1}{s+1/2} \right) + \frac{1}{s-1/2} \right),$$

using that $\gamma_s^+(\mathbf{t}_{v_{\ell,1}E}) = q^{s-1/2} 2(2\pi)^{1/2-s} \Gamma(s+1/2)$ and the formula (3.5) for additive twist L -series of oldforms. Now when considering the sum over $a \in (\mathbb{Z}/q\mathbb{Z})^\times$ twisted by $\chi \bmod q$ non-principal, we see that the poles at $s = 1/2$ cancel by orthogonality of characters. Thus, by letting $s \rightarrow 1/2$ in the Birch–Stevens formula (Proposition 5.1), we get

$$\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \overline{\chi(a)} \langle [\mathcal{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}} \\ = \frac{12i}{\pi} \frac{\nu(\chi^*, q^*; N) - \nu(\chi^*, q^*; 1)}{N-1} \tau(\bar{\chi}^*) L\left(\frac{1}{2}, E, \chi^*\right) \\ + \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \overline{\chi(a)} \frac{\text{tr}(\psi(a))}{q}, \quad (6.19)$$

where $q^* = q/\mathbf{c}(\chi)$. By Lemma 6.10, we arrive at the following expression for the first term in (6.19) valid in all cases:

$$\frac{-12i}{\pi} \frac{|1 - \chi^*(N)|^2}{N-1} \nu\left(\chi^*, \frac{q^*}{(q^*, N^\infty)}; 1\right) \tau(\bar{\chi}^*) L\left(\frac{1}{2}, E, \chi^*\right),$$

using here that $N|q$, which means that if $(N, q^*) = 1$, then $N|\mathbf{c}(\chi)$ (so that $\chi^*(N) = 0$).

We can express the Hecke eigenvalues of E as the convolution product

$$|\cdot|^{1/2} * |\cdot|^{-1/2},$$

which implies that

$$\tau(\bar{\chi}^*)L\left(\frac{1}{2}, E, \chi^*\right) = \tau(\bar{\chi}^*)L(1, \chi^*)L(0, \chi^*),$$

vanishing if χ^* is even. For χ^* odd we get by the functional equation for Dirichlet L -functions [12, Chapter 9, eq. (14)] that

$$\tau(\bar{\chi}^*)L\left(\frac{1}{2}, E, \chi^*\right) = \frac{ic(\chi)}{\pi}|L(1, \chi^*)|^2.$$

Putting all of this together yields the required result. $\blacksquare$

We are now ready to prove our main lower bound for the Eisenstein contribution, which is even uniform in the level N .

Corollary 6.11. *For each $\varepsilon > 0$, there are constants $c_1(\varepsilon), c_2(\varepsilon) > 0$ such that the following holds: Let $N \geq 2$ be a prime and $q \geq 2$ an integer such that $N|q$. Consider a double coset embedding ψ of level q and a subgroup $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$ satisfying*

- (1) $-1 \notin H$,
- (2) $\{a \in (\mathbb{Z}/q\mathbb{Z})^\times : a \equiv 1 \pmod{q/N}\} \not\subset H$.

Then, we have

$$\begin{aligned} \sum_{a \in H} \langle [\mathcal{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}} &\geq c_1(\varepsilon) \frac{q^{1-\varepsilon}}{N} - c_2(\varepsilon) |H|^{3/4} q^{1/6+\varepsilon} \\ &\quad + \sum_{a \in H} n_\psi(a) \end{aligned} \quad (6.20)$$

for q large enough.

Proof. By character orthogonality and Proposition 6.8, we have

$$\begin{aligned} &\sum_{a \in H} \langle [\mathcal{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}} \\ &= \frac{|H|}{\varphi(q)} \sum_{\substack{\chi \bmod q \\ \chi|_H=1}} \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} \overline{\chi(a)} \langle [\mathcal{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}} \\ &= \sum_{a \in H} \frac{\text{tr}(\psi(a))}{q} + \frac{12|H|}{\pi^2 \varphi(q)} \\ &\quad \times \sum_{\substack{\chi \bmod q \\ \chi|_H=1 \\ \chi(-1)=-1}} \frac{|\chi^*(N) - 1|^2 \mathbf{c}(\chi)}{N-1} |L(1, \chi^*)|^2 v\left(\chi^*, \frac{q^*}{(q^*, N^\infty)}; 1\right). \end{aligned} \quad (6.21)$$

To deal with the first term, we observe that by Lemma 2.3 and Corollary 6.5, we have

$$\sum_{a \in H} \frac{\text{tr}(\psi(a))}{q} = \sum_{a \in H} n_\psi(a) + \sum_{a \in H} \frac{t_a}{q} = \sum_{a \in H} n_\psi(a) + O_\varepsilon(|H|^{3/4} q^{1/6+\varepsilon}).$$

To give a lower bound for the sum over characters in (6.21) (note that all the summands are non-negative), we consider the following subset:

$$\mathcal{M} := \{\chi \bmod q : (q^*, N) = 1, \chi|_H = 1, \chi(-1) = -1\},$$

where $q^* = q/\mathbf{c}(\chi)$. We see by the principle of inclusion and exclusion,

$$|\mathcal{M}| = \frac{\varphi(q)}{|H|} - \frac{\varphi(q/N)}{|H \bmod q/N|} - \frac{\varphi(q)}{|H'|} + \frac{\varphi(q/N)}{|H' \bmod q/N|},$$

where $H' := \langle H, -1 \rangle \leq (\mathbb{Z}/q\mathbb{Z})^\times$. Note that $|H'| = 2|H|$ by the assumption $-1 \notin H$. By simple group theory, we see that

$$\frac{\varphi(q/N)}{|H \bmod q/N|} = \frac{\varphi(q)}{|H|} \frac{|\{h \in H : h \equiv 1 \bmod q/N\}|}{|\{a \in (\mathbb{Z}/q\mathbb{Z})^\times : a \equiv 1 \bmod q/N\}|},$$

and furthermore the above is an integer dividing $\varphi(q)/|H|$ (and similarly for H'). This means that

$$\begin{aligned} |\mathcal{M}| &= \frac{\varphi(q)}{2|H|} \left(1 - \frac{2|\{h \in H : h \equiv 1 \bmod q/N\}|}{|\{a \in (\mathbb{Z}/q\mathbb{Z})^\times : a \equiv 1 \bmod q/N\}|} \right. \\ &\quad \left. + \frac{|\{h \in H' : h \equiv 1 \bmod q/N\}|}{|\{a \in (\mathbb{Z}/q\mathbb{Z})^\times : a \equiv 1 \bmod q/N\}|} \right) \\ &\geq \frac{\varphi(q)}{2|H|} \left(1 - \frac{|\{h \in H : h \equiv 1 \bmod q/N\}|}{|\{a \in (\mathbb{Z}/q\mathbb{Z})^\times : a \equiv 1 \bmod q/N\}|} \right) \geq \frac{\varphi(q)}{4|H|}, \end{aligned}$$

where we have used the second assumption on H in the last inequality. Note that $(N, q^*) = 1$ implies $N|\mathbf{c}(\chi)$, and thus $\chi^*(N) = 0$. This means that for each $\chi \in \mathcal{M}$, we get a positive contribution in (6.21) of

$$\begin{aligned} &\frac{|\chi^*(N) - 1|^2 \mathbf{c}(\chi)}{N - 1} |L(1, \chi^*)|^2 v\left(\chi^*, \frac{q^*}{(q^*, N^\infty)}; 1\right) \\ &= \frac{\mathbf{c}(\chi)}{N - 1} |L(1, \chi^*)|^2 v(\chi^*, q^*; 1) \gg_\varepsilon \frac{q^{1-\varepsilon}}{N}, \end{aligned}$$

by Lemma 6.9 and the lower bounds $L(1, \chi^*) \gg_\varepsilon \mathbf{c}(\chi)^{-\varepsilon}$ and $\varphi(n) \gg_\varepsilon n^{1-\varepsilon}$. Combining this with the lower bound on $|\mathcal{M}|$ as above yields (6.20), as required. ■

6.3.1. In the case where $-1 \in H$. Similarly, we obtain the following lower bound in the case where $-1 \in H$.

Corollary 6.12. *Let $N \geq 2$ be prime. Let $q \geq 2$ be such that $N|q$ and let $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$ be a subgroup containing -1 . Then, for any coset $cH \subset (\mathbb{Z}/q\mathbb{Z})^\times$ and $\varepsilon > 0$, we have*

$$\sum_{a \in cH} \langle [\mathcal{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}} = \sum_{a \in cH} n_\psi(a) + O_\varepsilon(|H|^{3/4} q^{1/6+\varepsilon}).$$

Proof. By the assumption that $-1 \in H$, we have by Proposition 6.8 that

$$\sum_{a \in cH} \langle [\mathcal{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}} = \sum_{a \in cH} \frac{\text{tr}(\psi(a))}{q},$$

and thus the result follows by Lemma 2.3 and the bound (6.8) of Corollary 6.5, as above. $\blacksquare$

7. Proof of equidistribution

In this section, we will prove our main results using the above. We begin by some technical reductions which are necessary for the proof of Theorem 1.1.

7.1. Reduction to regular ψ

Let $N \geq 1$ and let $q \geq 2$ be such that $N|q$. Let ψ be a double coset embedding of level q and $cH \subset (\mathbb{Z}/q\mathbb{Z})^\times$ a coset. Then we want to prove that for any smooth, compactly supported function $\phi: \mathbf{T}^1(Y_0(N)) \rightarrow \mathbb{C}$, we have (with an explicit error term)

$$\frac{\sum_{a \in cH} \int_{\mathcal{C}_a^\psi(N)} \phi(t) d\mu_a^\psi(t)}{\sum_{a \in cH} |\mathcal{C}_a^\psi(N)|} \rightarrow \int_{\mathbf{T}^1(Y_0(N))} \phi(g) dg \quad \text{as } q \rightarrow \infty.$$

Special care will have to be given to the $a \in (\mathbb{Z}/q\mathbb{Z})^\times$ for which $\psi(a)$ is *singular* in the sense that $|\text{tr}(\psi(a))| < q^{1-\varepsilon}$ for some (small) $\varepsilon > 0$. Given a double coset embedding $\psi: (\mathbb{Z}/q\mathbb{Z})^\times \hookrightarrow \Gamma_0(N)$ of level q , we define for $\varepsilon > 0$ another double coset embedding of level q by

$$\psi_\varepsilon(a) := \begin{cases} \psi(a), & |\text{tr}(\psi(a))| > q^{1-\varepsilon}, \\ \psi(a) \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, & |\text{tr}(\psi(a))| \leq q^{1-\varepsilon}. \end{cases}$$

We say that ψ is ε -regular if $\psi = \psi_\varepsilon$ and notice that ψ_ε is itself ε -regular for q large enough since

$$\left| \operatorname{tr}(\psi(a) \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}) \right| = |q + \operatorname{tr}(\psi(a))| \geq q - |\operatorname{tr}(\psi(a))|.$$

We have the following result which allows us to reduce to the regular case.

Lemma 7.1. *Fix $\delta > 2/3$. Let $N \geq 1$ and $q \geq 2$ be integers such that $N|q$. Consider a double coset embedding ψ of level q and a coset $cH \subset (\mathbb{Z}/q\mathbb{Z})^\times$ of a subgroup $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$ of size $\geq q^\delta$. Then for any bounded function $\phi: \Gamma_0(N) \backslash \operatorname{PSL}_2(\mathbb{R}) \rightarrow \mathbb{C}$ and $\varepsilon > 0$, we have that*

$$\left| \frac{\sum_{a \in cH} \int_{\mathcal{C}_a^\psi(N)} \phi(t) d\mu_a^\psi(t)}{\sum_{a \in cH} |\mathcal{C}_a^\psi(N)|} - \frac{\sum_{a \in cH} \int_{\mathcal{C}_a^{\psi_\varepsilon}(N)} \phi(t) d\mu_a^{\psi_\varepsilon}(t)}{\sum_{a \in cH} |\mathcal{C}_a^{\psi_\varepsilon}(N)|} \right| \ll_\varepsilon \|\phi\|_\infty \left(q^{-\varepsilon} + \frac{q^{1/2+3\varepsilon}}{|H|^{3/4}} \right). \quad (7.1)$$

Proof. First of all, we see that

$$\begin{aligned} & \left| \frac{\sum_{a \in cH} (\int_{\mathcal{C}_a^\psi(N)} \phi(t) d\mu_a^\psi(t) - \int_{\mathcal{C}_a^{\psi_\varepsilon}(N)} \phi(t) d\mu_a^{\psi_\varepsilon}(t))}{\sum_{a \in cH} |\mathcal{C}_a^\psi(N)|} \right| \\ & \ll \|\phi\|_\infty \frac{\sum_{a \in cH: |\operatorname{tr}(\psi(a))| \leq q^{1-\varepsilon}} |\mathcal{C}_a^\psi(N)| + |\mathcal{C}_a^{\psi_\varepsilon}(N)|}{\sum_{a \in cH} |\mathcal{C}_a^\psi(N)|} \\ & \ll \|\phi\|_\infty \frac{\log q \cdot |\{a \in cH : |\operatorname{tr}(\psi(a))| \leq q^{1-\varepsilon}\}|}{\sum_{a \in cH} |\mathcal{C}_a^\psi(N)|}, \end{aligned} \quad (7.2)$$

since

$$|\operatorname{tr}(\psi_\varepsilon(a))| \leq q + |\operatorname{tr}(\psi(a))| \ll q$$

in the sum on the right-hand side of (7.2). Now by the bound (6.6) from Corollary 6.5 (with $T = q^\varepsilon$ and $\delta = 0$), we see that

$$\begin{aligned} |\{a \in cH : |\operatorname{tr}(\psi(a))| \leq q^{1-\varepsilon}\}| & \leq \left| \left\{ a \in cH : \frac{|ta|}{q} \leq q^{-\varepsilon} \right\} \right| \\ & \ll_\varepsilon |H| q^{-\varepsilon} + q^{1/2+3\varepsilon} |H|^{1/4}, \end{aligned}$$

and by Proposition 6.6, we have

$$\sum_{a \in cH} |\mathcal{C}_a^\psi(N)| \gg |H| \log q,$$

which shows that the left-hand side of (7.2) is bounded by the right-hand side of (7.1).

Similarly, we conclude that

$$\left| \sum_{a \in cH} \int_{\mathcal{C}_a^{\psi_\varepsilon}(N)} \phi(t) d\mu_a^{\psi_\varepsilon}(t) \left(\frac{1}{\sum_{a \in cH} |\mathcal{C}_a^\psi(N)|} - \frac{1}{\sum_{a \in cH} |\mathcal{C}_a^{\psi_\varepsilon}(N)|} \right) \right| \\ \ll \|\phi\|_\infty \left(q^{-\varepsilon} + \frac{q^{1/2+3\varepsilon}}{|H|^{3/4}} \right)$$

by Corollary 6.5, in view of Lemma 2.3, and Proposition 6.6. This gives the required result by the triangle inequality. ■

This means that we may assume that the double coset embedding ψ is ε -regular for some $\varepsilon > 0$, as long as $|H| \geq q^\delta$ for $\delta > 2/3$. We are now ready to prove our main result.

Theorem 7.2. *Fix $N \geq 1$, $0 < \kappa < 2$ and $\delta > 7/8$. For every $q \geq 2$ such that $N|q$, consider a coset $cH \subset (\mathbb{Z}/q\mathbb{Z})^\times$ with $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$ a subgroup of size $\geq q^\delta$ and a double coset embedding $\psi: (\mathbb{Z}/q\mathbb{Z})^\times \hookrightarrow \Gamma_0(N)$ of level q such that*

$$|\mathrm{tr}(\psi(a))| \leq q(\log q)^\kappa$$

for all $a \in cH$. Then, for any smooth and compactly supported function

$$\phi: \mathbf{T}^1(Y_0(N)) \rightarrow \mathbb{C} \quad \text{and} \quad \varepsilon > 0,$$

we have that

$$\frac{\sum_{a \in cH} \int_{\mathcal{C}_a^{\psi_\varepsilon}(N)} \phi(t) d\mu_a^{\psi_\varepsilon}(t)}{\sum_{a \in cH} |\mathcal{C}_a^\psi(N)|} = \int_{\mathbf{T}^1(Y_0(N))} \phi(g) dg + O_{\phi, \varepsilon}((\log q)^{\kappa/2-1+\varepsilon}).$$

Proof. By Weyl's criterion as in Lemma 2.4 and by Proposition 6.6, it suffices to prove the following bound for the Weyl sums:

$$\sum_{a \in cH} \int_{\mathcal{C}_a^\psi(N)} F(t) d\mu_a^\psi(t) \stackrel{?}{\ll}_\varepsilon |\mathbf{t}_F|^A (q^{7/8+\varepsilon} + |H|(\log q)^{\kappa/2+\varepsilon}),$$

for $F \in \mathcal{B}_0(\Gamma_0(N)) \cup \mathcal{B}_\varepsilon(\Gamma_0(N))$ and some (fixed) $A > 0$. By applying Theorem 5.2, the problem is reduced to bounding the remaining terms depending on the traces of ψ on the right-hand sides of (5.3) and (5.4). By the assumption that $|\mathrm{tr}(\psi(a))| \leq q(\log q)^\kappa$, we see that

$$\left(\frac{|\mathrm{tr}(\psi(a))|}{q} \right)^{1/2+\varepsilon} \leq (\log q)^{\kappa/2+2\varepsilon},$$

as required. By Lemma 7.1 and the assumption that $|H| \geq q^\delta$ with $\delta > 7/8$, we may assume that ψ is ε -regular for some $0 < \varepsilon < 1/100$, and thus by an application of the estimate (6.6) from Corollary 6.5, we get

$$\begin{aligned} \sum_{a \in cH} \left(\frac{q}{|\operatorname{tr}(\psi(a))|} \right)^{1/2+\varepsilon} &\leq |\{a \in cH : |\operatorname{tr}(\psi(a))| \geq q\}| \\ &\quad + \sum_{1 \leq m \leq \varepsilon \log q} \sum_{\substack{a \in cH: \\ q/2^m \leq |\operatorname{tr}(\psi(a))| < q/2^{m-1}}} 2^{m(1/2+\varepsilon)} \\ &\ll |H| + |H| \sum_{1 \leq m \leq \varepsilon \log q} 2^{-m} 2^{m(1/2+\varepsilon)} + q^{1/2+3\varepsilon} |H|^{1/4} \\ &\ll_\varepsilon |H|. \end{aligned}$$

Thus, we conclude the required effective convergence using Weyl's criterion as in Lemma 2.4. ■

Remark 7.3. Notice that when $N = 1$, $q = p$ is prime and $H = (\mathbb{Z}/p\mathbb{Z})^\times$, the Weyl sum of a Hecke–Maaß form f is of the shape

$$c_f L\left(\frac{1}{2}, f\right) v(f, \chi_0, p; 1) = c_f L\left(\frac{1}{2}, f\right) (p^{1/2} \lambda_f(p) - 2),$$

for some constant c_f , plus an error-term that is negligible compared to the main term (i.e. the last terms on the right-hand side of (5.3) and (5.4)). Thus equidistribution as $p \rightarrow \infty$ in this case follows from any power saving bound for the Hecke eigenvalues. This is exactly because the Weyl sums in this case are essentially obtained by integrating the automorphic forms along the Hecke orbits $\{\gamma_t i \mathbb{R} : t \in \mathbf{P}^1(\mathbb{F}_p)\}$ with notation as in (1.3) above.

7.2. Concentration in homology

Recall that for a finite-dimensional vector space V we defined the *sphere associated to V* as $\mathbf{S}(V) := (V - \{0\})/\mathbb{R}_{>0}$ equipped with the quotient topology and denoted by $v \mapsto \bar{v}$ the canonical projection for $v \neq 0$. By Weyl's criterion as in Lemma 2.6, the convergence

$$\bar{v}_n \rightarrow \bar{v}_0 \quad \text{in } \mathbf{S}(V),$$

is equivalent to showing, for some (fixed) norm $\|\cdot\|: V \rightarrow \mathbb{R}_{\geq 0}$, that

$$\left\| \frac{v_n}{\|v_n\|} - \frac{v_0}{\|v_0\|} \right\| \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (7.3)$$

In our setting, we consider the vector space $V_N = H_1(Y_0(N), \mathbb{R})$ with N prime. We will actually obtain bounds for the left-hand side of (7.3) *uniform* in terms of N for

certain natural norms on V_N (this corresponds to the *level aspect* in our equidistribution problem cf. [36]). Given a basis B^* of the dual $V_N^* \cong H^1(Y_0(N), \mathbb{R})$, we define the following norm on V_N :

$$\|v\|_{B^*} := \max_{\omega \in B^*} |\langle v, \omega \rangle_{\text{cap}}|. \quad (7.4)$$

It is a standard fact [51, eq. (3)] that the following matrices generate $\Gamma_0(N)$:

$$\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} a & -(aa^* + 1)/N \\ N & -a^* \end{pmatrix} : 0 < a < N \right\}, \quad (7.5)$$

where $0 < a^* < N$ is such that $aa^* \equiv -1 \pmod{N}$. For each prime N , consider a basis B of $H_1(Y_0(N), \mathbb{R})$ containing $v_E(N)$ and consisting of images of elements of (7.5) under the isomorphism

$$H_1(Y_0(N), \mathbb{Z}) \cong \Gamma_0(N)^{\text{ab}} / (\Gamma_0(N)^{\text{ab}})_{\text{tor}}.$$

Let B^* be the dual basis and $\|\cdot\|_N := \|\cdot\|_{B^*}$ the associated norm on $H_1(Y_0(N), \mathbb{R})$, as defined in (7.4). The system of norms $(\|\cdot\|_N)_{N \text{ prime}}$ will serve as a set of “natural norms” when varying N . They have a nice interpretation in terms of group theory, as we will see.

Theorem 7.4. *Let $N \geq 2$ be a prime, $0 < \varepsilon < 1/100$ and let $\|\cdot\|_N$ be a norm on $H_1(Y_0(N), \mathbb{R})$ defined as above. For $q \geq 2$ such that $N|q$, let $H \subset (\mathbb{Z}/q\mathbb{Z})^\times$ be a subgroup and let ψ a double coset embedding of level q such that*

- (1) $-1 \notin H$,
- (2) $\{a \in (\mathbb{Z}/q\mathbb{Z})^\times : a \equiv 1 \pmod{q/N}\} \not\subset H$,
- (3) $|\sum_{a \in H} n_\psi(a)| \leq q^{1-\varepsilon}/N$.

Then, we have

$$\left\| \frac{\sum_{a \in H} [\mathcal{C}_a^\psi(N)]}{\|\sum_{a \in H} [\mathcal{C}_a^\psi(N)]\|_N} - \frac{v_E(N)}{\|v_E(N)\|_N} \right\|_N \ll_\varepsilon N^{9/4} q^{-1/8+\varepsilon}. \quad (7.6)$$

Proof. We start by noting that the left-hand side of (7.6) is trivially bounded by 2 by the triangle inequality. Thus we may assume that $q \geq N^{18}$. Let B be a basis for V_N associated to a subset of (7.5) containing $v_E(N)$ and let $B^* \subset H^1(Y_0(N), \mathbb{R})$ be the dual basis with respect to the cap product pairing. By expanding in the Hecke basis (2.38), we have for $\omega \in B^*$ that

$$\begin{aligned} \sum_{a \in H} \langle [\mathcal{C}_a^\psi(N)], \omega \rangle_{\text{cap}} &= \sum_{f \in \mathcal{B}_N, \pm} \langle v_f^\pm, \omega \rangle_{\text{cap}} \sum_{a \in H} \langle [\mathcal{C}_a^\psi(N)], \omega_f^\pm \rangle_{\text{cap}} \\ &\quad + \langle v_E(N), \omega \rangle_{\text{cap}} \sum_{a \in H} \langle [\mathcal{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}}. \end{aligned} \quad (7.7)$$

Now that by the definition of the cap product pairing and of the classes $\omega_f^\pm$ in (2.39), we have

$$\langle [\mathcal{C}_a^\psi(N)], \omega_f^\pm \rangle_{\text{cap}} = \frac{1}{1+i \pm (1-i)} \left(2\pi i \int_{\mathcal{C}_a^\psi(N)} f(z) dz \pm 2\pi i \overline{\int_{\mathcal{C}_a^\psi(N)} f(z) dz} \right).$$

Thus by inserting the bound from Corollary 5.3 (note that the first term is dominating since $q \geq N^{18}$), we conclude that the cuspidal contribution in (7.7) is bounded by

$$\ll_\varepsilon N^{1/4} q^{7/8+\varepsilon/2} \sum_{f \in \mathcal{B}_N, \pm} |\langle v_f^\pm, \omega \rangle_{\text{cap}}| \ll_\varepsilon N^{5/4} q^{7/8+\varepsilon},$$

where we in the last inequality applied Cauchy–Schwarz and the L^2 -bound for cap product pairings from [40, Theorem 6.1]. Now, by the lower bound from Corollary 6.11 (applied with $\varepsilon/2$) and assumption (3), we conclude for q sufficiently large that

$$\begin{aligned} \sum_{a \in H} \langle [\mathcal{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}} &\geq c_1 \left(\frac{\varepsilon}{2} \right) \frac{q^{1-\varepsilon/2}}{N} - c_2(\varepsilon/2) |H|^{3/4} q^{1/6+\varepsilon/2} - \frac{q^{1-\varepsilon}}{N} \\ &\gg_\varepsilon \frac{q^{1-\varepsilon/2}}{N}, \end{aligned}$$

using also $N \leq q^{1/18}$ and $\varepsilon < 1/100$, so that

$$|H|^{3/4} q^{1/6+\varepsilon/2} \leq q^{11/12+1/200} \leq q^{1-1/100} q^{-1/18} \leq q^{1-\varepsilon}/N.$$

Now by applying the effective Weyl’s criterion as in Lemma 2.6 to the vectors

$$\frac{\sum_{a \in H} [\mathcal{C}_a^\psi(N)]}{\sum_{a \in H} \langle [\mathcal{C}_a^\psi(N)], \omega_E(N) \rangle_{\text{cap}}},$$

we conclude the required bound (7.6). ■

Now if we keep N fixed we arrive at the following concentration result by Weyl’s criterion as in equation (2.44) of Lemma 2.6.

Corollary 7.5. *Fix a prime $N \geq 3$ and $\varepsilon > 0$. For each positive integer $q \geq 2$ with $N|q$, consider a subgroup $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$ and a double coset embedding ψ of level q such that*

- (1) $-1 \notin H$,
- (2) $\{a \in (\mathbb{Z}/q\mathbb{Z})^\times : a \equiv 1 \pmod{q/N}\} \not\subset H$,
- (3) $|\sum_{a \in H} n_\psi(a)| \leq q^{1-\varepsilon}$.

Then, as $q \rightarrow \infty$, we have that

$$\overline{\sum_{a \in H} [\mathcal{C}_a^\psi(N)]} \rightarrow \overline{v_E(N)} \quad \text{in } \mathbf{S}(V_N).$$

Finally, let us remark that condition (3) in the corollary above is implied by condition (3) in the introduction, saying that the absolute trace is minimal: to see this note that in view of Lemma 2.3 the trace being minimal corresponds to $n_\psi(a) = 0$ for all $a \in H$ with $a + \bar{a} \notin \{0, \pm 1, \pm 2\} \bmod q$, and $n_\psi(a) \in \{\pm 1\}$ otherwise, and we have for any $\varepsilon > 0$ the bound

$$\begin{aligned} & |\{a \in (\mathbb{Z}/q\mathbb{Z})^\times : a + \bar{a} \in \{0, \pm 1, \pm 2\} \bmod q\}| \\ &= |\{a \in (\mathbb{Z}/q\mathbb{Z})^\times : \exists k \in \{0, \pm 1, \pm 2\} \text{ s.t. } q|(a^2 + ka + 1)\}| \\ &\ll_\varepsilon q^{1/2+\varepsilon} \quad \text{as } q \rightarrow \infty. \end{aligned} \tag{7.8}$$

In conclusion, Corollary 7.5 implies Theorem 1.3.

Remark 7.6. The natural bases of $H_1(Y_0(N), \mathbb{R})$ constructed above are analogues of the set of isomorphism classes of supersingular elliptic curves $\mathcal{E}\ell\ell^{ss}(\mathbb{F}_{p^2})$ defined over $\mathbb{F}_{p^2}$ considered in the context of supersingular reduction of CM-elliptic curves. For details on this analogy, see [40, Section 2.1].

7.2.1. The case $-1 \in H$. In this section, we will obtain results in the complementary case where the subgroup H contains -1 . In this case, we will have to require the traces of the double coset embeddings to be biased.

Theorem 7.7. *Let $N \geq 2$ be prime and $\delta > 1/6$, and let $\|\cdot\|_N$ be a norm on $H_1(Y_0(N), \mathbb{R})$ defined above. For $q \geq 2$ such that $N|q$, let $H \subset (\mathbb{Z}/q\mathbb{Z})^\times$ be a subgroup containing -1 and let ψ be a double coset embedding of level q such that*

$$\sum_{a \in cH} n_\psi(a) \geq M \geq |H|^{3/4} q^\delta. \tag{7.9}$$

Then, for each $\varepsilon > 0$, we have that

$$\left\| \frac{\sum_{a \in H} [\mathcal{C}_a^\psi(N)]}{\|\sum_{a \in H} [\mathcal{C}_a^\psi(N)]\|_N} - \frac{v_E(N)}{\|v_E(N)\|_N} \right\|_N \ll_\varepsilon \frac{N^{5/4} q^{7/8+\varepsilon} + N^{7/4} q^{3/4+\varepsilon}}{M}. \tag{7.10}$$

Proof. The proof follows from Corollary 5.3 exactly as above using now Corollary 6.12 and the assumption (7.9) to lower bound the contribution from the Eisenstein class $\omega_E(N)$. ■

Again we conclude the following by Weyl's criterion as in Lemma 2.6.

Corollary 7.8. Fix a prime $N \geq 2$ and $\varepsilon > 0$. For each $q \geq 2$ with $N|q$, consider a double coset embedding ψ of level q and a coset $cH \subset (\mathbb{Z}/q\mathbb{Z})^\times$ of a subgroup H containing -1 such that

$$\sum_{a \in cH} n_\psi(a) \geq |H|^{3/4} q^{1/6+\varepsilon} + q^{7/8+\varepsilon}.$$

Then, as $q \rightarrow \infty$, we have that

$$\overline{\sum_{a \in cH} [\mathcal{C}_a^\psi(N)]} \rightarrow \overline{v_E(N)} \quad \text{in } \mathbf{S}(V_N).$$

7.2.2. An application to group theory. Finally, we will give a group theoretic application of Theorems 7.4 and 7.7. For this part, we will assume that N is prime and $N \equiv -1 \pmod{12}$. Recall from the introduction that this means that $\Gamma_0(N)$ is torsion-free [29, eqs. (2.12), (2.13)], and thus $\Gamma_0(N)$ is a free group on $2g + 1$ generators, where $g = N/12 + O(1)$. In this case, Doan–Kim–Lang–Tan [13] have recently constructed a subset of the generators (1.5) containing the matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, which defines a set of free generators of $\Gamma_0(N)$. We will be interested in the representation of matrices in this free basis which in general is a very hard question.

Proof of Corollary 1.4. Assume for now that q is square-free. Let ψ be the double coset embedding of level q defined by

$$\psi(a) = \begin{pmatrix} a & -\frac{aa^*+1}{q} \\ q & -a^* \end{pmatrix} \quad (7.11)$$

for $0 < a, a^* < q$ such that $aa^* \equiv -1 \pmod{q}$. Notice that the above matrices are indeed hyperbolic by the assumptions: clearly $\psi(a) \in \Gamma_0(N)$, which is torsion-free by the assumption that $N \equiv -1 \pmod{12}$, and so the only way hyperbolicity could fail is that the trace is ± 2 , which implies that $a(\pm 2 - a) \equiv 1 \pmod{q} \Leftrightarrow q|(a \mp 1)^2$, and so $a \in \{1, q-1\}$ since q is assumed square-free. In this last case, the matrix is explicitly checked to be hyperbolic by the assumption $q \geq N \geq 11$. Recall that by (6.8) of Corollary 6.5, we have

$$\sum_{a \in H} \frac{\text{tr}(\psi(a))}{q} = \sum_{a \in H} n_\psi(a) + \sum_{a \in H} \frac{t_a}{q} = \sum_{a \in H} n_\psi(a) + O_\varepsilon(|H|^{3/4} q^{1/6+\varepsilon}). \quad (7.12)$$

On the other hand, by (6.9) of Corollary 6.5, we conclude that

$$\begin{aligned} \sum_{a \in H} \frac{\text{tr}(\psi(a))}{q} &= \left(\sum_{\substack{0 < a < q: \\ (a \bmod q) \in H}} \frac{a}{q} \right) - \left(\sum_{\substack{0 < a < q: \\ (a \bmod q) \in -H}} \frac{a}{q} \right) \\ &= |H| \left(\frac{1}{2} - \frac{1}{2} \right) + O_\varepsilon(|H|^{2/3} q^{1/6+\varepsilon}). \end{aligned} \quad (7.13)$$

This means that

$$\sum_{a \in H} n_\psi(a) \ll_\varepsilon |H|^{2/3} q^{1/6+\varepsilon} + |H|^{3/4} q^{1/6+\varepsilon} \ll_\varepsilon q^{11/12+\varepsilon},$$

and since $N \ll_\delta q^{1/\delta}$ with $\delta > 18$, condition (3) of Theorem 7.4 is satisfied for q large enough and $\varepsilon > 0$ small enough. If we assume that a conjugate of the matrix (1.6) is in fact contained in the subgroup generated by $\sigma_1, \dots, \sigma_{2g}$, then we get

$$\left\| \frac{\sum_{a \in H} [\mathcal{C}_a^\psi(N)]}{\|\sum_{a \in H} [\mathcal{C}_a^\psi(N)]\|_{B^*}} - \frac{v_E(N)}{\|v_E(N)\|_{B^*}} \right\|_{B^*} \geq 1,$$

where B^* is the dual basis of the basis of $H_1(Y_0(N), \mathbb{R})$ corresponding to

$$\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \sigma_1, \dots, \sigma_{2g} \right\}.$$

This contradicts the bound (7.6) of Theorem 7.4 by the assumption that $q \geq c(\delta)N^\delta$ with $\delta > 18$. If q is not square-free, then the matrix (7.11) is not hyperbolic whenever $a \neq \pm 1$ and $q|(a \pm 1)^2$. In these cases, we multiply the matrix (7.11) by $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ from the right to obtain a hyperbolic matrix and consider this double coset embedding instead. This has a negligible effect on the sums (7.12) and (7.13) by the estimate (7.8), and so we conclude in this case as well. ■

Similarly, we get the following corollary in the case where $-1 \in H$, in which we need to consider matrices associated to $a \in H$ with *positive* traces (meaning that the associated double coset embeddings are biased), so that Theorem 7.7 applies.

Corollary 7.9. *Let $N \equiv -1 \pmod{12}$ be prime and let*

$$\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \sigma_1, \dots, \sigma_{2g} \right\},$$

be a subset of (7.5), which defines a set of free generators of $\Gamma_0(N)$. Then for each $\delta > 7/8$, there exists a constant $c(\delta) > 0$ such that the following holds: Let $q \geq N^4$ be an integer such that $N|q$ and let $0 < a_1, \dots, a_m < q$ with $m \geq c(\delta)N^{5/4}q^\delta$ be pairwise distinct integers coprime to q such that

$$\{a_i \bmod q : 1 \leq i \leq m\} \subset (\mathbb{Z}/q\mathbb{Z})^\times,$$

is a coset of a subgroup containing -1 . For $1 \leq i \leq m$, let $0 < \bar{a}_i < q$ be defined by $q|a_i\bar{a}_i - 1$. Then no conjugate of

$$\begin{pmatrix} a_1 & \frac{a_1\bar{a}_1-1}{q} \\ q & \bar{a}_1 \end{pmatrix} \begin{pmatrix} a_2 & \frac{a_2\bar{a}_2-1}{q} \\ q & \bar{a}_2 \end{pmatrix} \dots \begin{pmatrix} a_m & \frac{a_m\bar{a}_m-1}{q} \\ q & \bar{a}_m \end{pmatrix} \in \Gamma_0(N)$$

is contained in the subgroup of $\Gamma_0(N)$ generated by the matrices $\sigma_1, \dots, \sigma_{2g}$.

Proof. Let

$$\{a_i \bmod q : 1 \leq i \leq m\} = cH,$$

with $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$ a subgroup and $c \in (\mathbb{Z}/q\mathbb{Z})^\times$, and define the double coset embedding

$$\psi(a) = \begin{pmatrix} a & \frac{a\bar{a}-1}{q} \\ q & \bar{a} \end{pmatrix}, \quad 1 < a < q-1, \quad \psi(\pm 1) = \begin{pmatrix} \pm 1 & 1 \\ q & \pm(q+1) \end{pmatrix}.$$

Note that indeed the matrices above are hyperbolic. By equation (6.9) of Corollary 6.5, we have

$$\begin{aligned} \sum_{a \in cH} \frac{\text{tr}(\psi(a))}{q} &= \left(\sum_{\substack{0 < a < q: \\ (a \bmod q) \in cH}} \frac{a}{q} \right) + \left(\sum_{\substack{0 < a < q: \\ (a \bmod q) \in \bar{c}H}} \frac{a}{q} \right) \\ &= |H| \left(\frac{1}{2} + \frac{1}{2} \right) + O_\varepsilon(|H|^{2/3} q^{1/6+\varepsilon}), \end{aligned}$$

which means in view of (7.12) that

$$\begin{aligned} \sum_{a \in cH} n_\psi(a) &= |H| (1 + O_\varepsilon(|H|^{-1/3} q^{1/6+\varepsilon} + |H|^{-1/4} q^{1/6+\varepsilon})) \\ &= |H| (1 + O_\varepsilon(q^{1/6-7/32+\varepsilon})) = |H| (1 + o(1)), \end{aligned}$$

by the assumption on H . This shows that condition (7.9) of Theorem 7.7 is satisfied, using here again that $|H| \gg q^{7/8}$, so that $|H| \gg |H|^{3/4} q^{1/6+\varepsilon}$ for ε sufficiently small. Now the result follows exactly as above using Theorem 7.7 in place of Theorem 7.4, where the right-hand side of (7.10) tends to zero since

$$\sum_{a \in cH} n_\psi(a) \gg |H| \gg_\delta N^{5/4} q^\delta \quad \text{and} \quad q \geq N^4. \quad \blacksquare$$

Acknowledgements. The author would like to thank Sary Drappeau, Daniel Kriz and Farrell Brumley for useful discussions, as well as Igor Shparlinski for suggesting the simplified proof of Corollary 6.2. The author would also like to give a special thanks to the referees for exceptionally thorough readings, insightful remarks and for pointing out many imprecisions. The comments helped to raise the quality of the paper significantly.

Funding. The author's research was supported by the Independent Research Fund Denmark DFF-1025-00020B.

References

- [1] M. Aka and M. Einsiedler, [Duke's theorem for subcollections](#). *Ergodic Theory Dynam. Systems* **36** (2016), no. 2, 335–342 Zbl [1355.37004](#) MR [3503026](#)
- [2] V. Blomer, F. Brumley, and I. Khayutin, The mixing conjecture under GRH. [v1] 2022, [v2] 2025, [arXiv:2212.06280v2](#)
- [3] V. Blomer and G. Harcos, [Hybrid bounds for twisted \$L\$ -functions](#). *J. Reine Angew. Math.* **621** (2008), 53–79 Zbl [1193.11044](#) MR [2431250](#)
- [4] V. Blomer and G. Harcos, [The spectral decomposition of shifted convolution sums](#). *Duke Math. J.* **144** (2008), no. 2, 321–339 Zbl [1246.11108](#) MR [2437682](#)
- [5] V. Blomer and P. Michel, [The unipotent mixing conjecture](#). *J. Anal. Math.* **151** (2023), no. 1, 25–57 Zbl [1543.11050](#) MR [4682945](#)
- [6] V. Blomer and D. Milićević, [The second moment of twisted modular \$L\$ -functions](#). *Geom. Funct. Anal.* **25** (2015), no. 2, 453–516 Zbl [1400.11097](#) MR [3334233](#)
- [7] J. Bourgain and A. Kontorovich, [Beyond expansion II: Low-lying fundamental geodesics](#). *J. Eur. Math. Soc. (JEMS)* **19** (2017), no. 5, 1331–1359 Zbl [1403.11054](#) MR [3635355](#)
- [8] J. Bourgain and S. V. Konyagin, [Estimates for the number of sums and products and for exponential sums over subgroups in fields of prime order](#). *C. R. Math. Acad. Sci. Paris* **337** (2003), no. 2, 75–80 Zbl [1041.11056](#) MR [1998834](#)
- [9] R. W. Bruggeman and Y. Motohashi, [A new approach to the spectral theory of the fourth moment of the Riemann zeta-function](#). *J. Reine Angew. Math.* **579** (2005), 75–114 Zbl [1064.11059](#) MR [2124019](#)
- [10] D. Bump, [Automorphic forms and representations](#). Cambridge Stud. Adv. Math. 55, Cambridge University Press, Cambridge, 1997 Zbl [0911.11022](#) MR [1431508](#)
- [11] H. Darmon, [Heegner points, Stark–Heegner points, and values of \$L\$ -series](#). In *International Congress of Mathematicians. Vol. II*, pp. 313–345, European Mathematical Society, Zürich, 2006 Zbl [1157.11023](#) MR [2275600](#)
- [12] H. Davenport, [Multiplicative number theory](#). Third edn., Grad. Texts in Math. 74, Springer, New York, 2000 Zbl [1002.11001](#) MR [1790423](#)
- [13] N. M. Doan, S. hyun Kim, M. L. Lang, and S. P. Tan, Optimal independent generating system for the congruence subgroups $\Gamma_0(p)$ and $\Gamma_0(p^2)$. [v1] 2022, [v3] 2023, [arXiv:2209.13937v3](#)
- [14] S. Drappeau and A. C. Nordentoft, Central values of additive twists of Maaß forms L -functions. [v1] 2022, [v2] 2026, [arXiv:2208.14346v2](#)
- [15] W. Duke, [Hyperbolic distribution problems and half-integral weight Maass forms](#). *Invent. Math.* **92** (1988), no. 1, 73–90 Zbl [0628.10029](#) MR [0931205](#)
- [16] W. Duke, J. B. Friedlander, and H. Iwaniec, [The subconvexity problem for Artin \$L\$ -functions](#). *Invent. Math.* **149** (2002), no. 3, 489–577 Zbl [1056.11072](#) MR [1923476](#)
- [17] W. Duke, Ö. Imamoğlu, and Á. Tóth, [Kronecker's first limit formula, revisited](#). *Res. Math. Sci.* **5** (2018), no. 2, article no. 20 Zbl [1441.11085](#) MR [3782448](#)
- [18] M. Einsiedler, E. Lindenstrauss, P. Michel, and A. Venkatesh, [Distribution of periodic torus orbits on homogeneous spaces](#). *Duke Math. J.* **148** (2009), no. 1, 119–174 Zbl [1172.37003](#) MR [2515103](#)

- [19] M. Einsiedler, E. Lindenstrauss, P. Michel, and A. Venkatesh, [The distribution of closed geodesics on the modular surface, and Duke's theorem](#). *Enseign. Math. (2)* **58** (2012), no. 3-4, 249–313 Zbl [1312.37032](#) MR [3058601](#)
- [20] V. Erlandsson and J. Souto, [Counting and equidistribution of reciprocal geodesics and dihedral groups](#). *Int. Math. Res. Not. IMRN* (2024), no. 13, 10298–10318 Zbl [1560.22031](#) MR [4770371](#)
- [21] É. Fouvry, E. Kowalski, and P. Michel, [Algebraic twists of modular forms and Hecke orbits](#). *Geom. Funct. Anal.* **25** (2015), no. 2, 580–657 Zbl [1344.11036](#) MR [3334236](#)
- [22] K. Golubev and A. Kamber, [On Sarnak's density conjecture and its applications](#). *Forum Math. Sigma* **11** (2023), article no. e48 Zbl [1548.11084](#) MR [4603107](#)
- [23] G. Harcos and P. Michel, [The subconvexity problem for Rankin–Selberg \$L\$ -functions and equidistribution of Heegner points. II](#). *Invent. Math.* **163** (2006), no. 3, 581–655 Zbl [1111.11027](#) MR [2207235](#)
- [24] A. Hatcher, *Algebraic topology*. Cambridge University Press, Cambridge, 2002 Zbl [1044.55001](#) MR [1867354](#)
- [25] J. Hoffstein and P. Lockhart, [Coefficients of Maass forms and the Siegel zero](#). With an appendix by Dorian Goldfeld, Jeffrey Hoffstein and Daniel Lieman. *Ann. of Math. (2)* **140** (1994), no. 1, 161–181 Zbl [0814.11032](#) MR [1289494](#)
- [26] P. Humphries, [Distributing points on the torus via modular inverses](#). *Q. J. Math.* **73** (2022), no. 1, 1–16 Zbl [1503.11106](#) MR [4395069](#)
- [27] P. Humphries and A. C. Nordentoft, [Sparse equidistribution of geometric invariants of real quadratic fields](#). *J. Eur. Math. Soc.* (2025), published online first
- [28] H. Iwaniec, [Fourier coefficients of modular forms of half-integral weight](#). *Invent. Math.* **87** (1987), no. 2, 385–401 Zbl [0606.10017](#) MR [0870736](#)
- [29] H. Iwaniec, [Spectral methods of automorphic forms](#). Second edn., Grad. Stud. Math. 53, American Mathematical Society, Providence, RI; Revista Matemática Iberoamericana, Madrid, 2002 Zbl [1006.11024](#) MR [1942691](#)
- [30] H. Iwaniec and E. Kowalski, [Analytic number theory](#). Amer. Math. Soc. Colloq. Publ. 53, American Mathematical Society, Providence, RI, 2004 Zbl [1059.11001](#) MR [2061214](#)
- [31] H. Iwaniec, W. Luo, and P. Sarnak, [Low lying zeros of families of \$L\$ -functions](#). *Inst. Hautes Études Sci. Publ. Math.* **91** (2000), 55–131 Zbl [1012.11041](#) MR [1828743](#)
- [32] I. Khayutin, [Joint equidistribution of CM points](#). *Ann. of Math. (2)* **189** (2019), no. 1, 145–276 Zbl [1431.11080](#) MR [3898709](#)
- [33] H. H. Kim, [Functoriality for the exterior square of \$GL_4\$ and the symmetric fourth of \$GL_2\$](#) . *J. Amer. Math. Soc.* **16** (2003), no. 1, 139–183 Zbl [1018.11024](#) MR [1937203](#)
- [34] A. Knightly and C. Li, [Kuznetsov's trace formula and the Hecke eigenvalues of Maass forms](#). *Mem. Amer. Math. Soc.* **224** (2013), no. 1055 Zbl [1314.11038](#) MR [3099744](#)
- [35] Y. V. Linnik, *Ergodic properties of algebraic fields*. *Ergeb. Math. Grenzgeb.* 45, Springer, New York, 1968 Zbl [0162.06801](#) MR [0238801](#)
- [36] S.-C. Liu, R. Masri, and M. P. Young, [Subconvexity and equidistribution of Heegner points in the level aspect](#). *Compos. Math.* **149** (2013), no. 7, 1150–1174 Zbl [1329.11046](#) MR [3078642](#)

- [37] S.-C. Liu, R. Masri, and M. P. Young, [Rankin–Selberg \$L\$ -functions and the reduction of CM elliptic curves](#). *Res. Math. Sci.* **2** (2015), article no. 22 Zbl [1380.11060](#) MR [3402822](#)
- [38] P. Michel and A. Venkatesh, [Equidistribution, \$L\$ -functions and ergodic theory: On some problems of Yu. Linnik](#). In *International Congress of Mathematicians. Vol. II*, pp. 421–457, European Mathematical Society, Zürich, 2006 Zbl [1157.11019](#) MR [2275604](#)
- [39] A. C. Nordentoft, [Central values of additive twists of cuspidal \$L\$ -functions](#). *J. Reine Angew. Math.* **776** (2021), 255–293 Zbl [1532.11062](#) MR [4279108](#)
- [40] A. C. Nordentoft, [Concentration of closed geodesics in the homology of modular curves](#). *Forum Math. Sigma* **11** (2023), article no. e91 Zbl [1540.11051](#) MR [4651628](#)
- [41] A. Ostafe, I. E. Shparlinski, and J. F. Voloch, [Equations and character sums with matrix powers, Kloosterman sums over small subgroups, and quantum ergodicity](#). *Int. Math. Res. Not. IMRN* (2023), no. 16, 14196–14238 Zbl [1537.81028](#) MR [4631431](#)
- [42] I. Petrow and M. P. Young, [The fourth moment of Dirichlet \$L\$ -functions along a coset and the Weyl bound](#). *Duke Math. J.* **172** (2023), no. 10, 1879–1960 Zbl [1544.11068](#) MR [4624371](#)
- [43] A. A. Popa, [Central values of Rankin \$L\$ -series over real quadratic fields](#). *Compos. Math.* **142** (2006), no. 4, 811–866 Zbl [1144.11041](#) MR [2249532](#)
- [44] P. Sarnak, [Class numbers of indefinite binary quadratic forms](#). *J. Number Theory* **15** (1982), no. 2, 229–247 Zbl [0499.10021](#) MR [0675187](#)
- [45] P. Sarnak, Reciprocal geodesics. in *Analytic number theory*, pp. 217–237, Clay Math. Proc. 7, American Mathematical Society, Providence, RI, 2007 Zbl [1198.11039](#) MR [2362203](#)
- [46] G. Shimura, [On the holomorphy of certain Dirichlet series](#). *Proc. London Math. Soc. (3)* **31** (1975), no. 1, 79–98 Zbl [0311.10029](#) MR [0382176](#)
- [47] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*. Publ. Math. Soc. Japan 11, Princeton University Press, Princeton, NJ, 1994 Zbl [0872.11023](#) MR [1291394](#)
- [48] B. F. Skubenko, The asymptotic distribution of integers on a hyperboloid of one sheet and ergodic theorems (Russian). *Izv. Akad. Nauk SSSR, Ser. Mat.* **26** (1962), no. 5, 721–752 Zbl [0107.04201](#) MR [0151436](#)
- [49] J.-L. Waldspurger, Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie. *Compositio Math.* **54** (1985), no. 2, 173–242 Zbl [0567.10021](#) MR [0783511](#)
- [50] M. P. Young, [Explicit calculations with Eisenstein series](#). *J. Number Theory* **199** (2019), 1–48 Zbl [1454.11083](#) MR [3926186](#)
- [51] D. Zagier, [Modular parametrizations of elliptic curves](#). *Canad. Math. Bull.* **28** (1985), no. 3, 372–384 Zbl [0579.14027](#) MR [0790959](#)
- [52] D. Zagier, Quantum modular forms. In *Quanta of maths*, pp. 659–675, Clay Math. Proc. 11, American Mathematical Society, Providence, RI, 2010 Zbl [1294.11084](#) MR [2757599](#)

- [53] S. Zelditch, [Selberg trace formulae and equidistribution theorems for closed geodesics and Laplace eigenfunctions: Finite area surfaces](#). *Mem. Amer. Math. Soc.* **96** (1992), no. 465 Zbl [0753.11023](#) MR [1101972](#)

Received 27 February 2024.

Asbjørn Christian Nordentoft

Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5,
2100 Copenhagen, Denmark; nordentoft@math.ku.dk

Author IDs: zbMATH [nordentoft.asbjorn-christian](#) MR [1416007](#)